محاضرات في الإحصاء الرياضي

إبراهيم محمد العلي   فتاة صبوح

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3} \\ \dots & \dots & \dots & \dots & .. \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{m} \end{bmatrix}$$



جامعة تشرين – كلية الاقتصاد

محاضرات

في

الاحصاء الرياضي

الدكتور

إبراهيم محمد العلي

الدكتورة

فتاة صبوح

لطلاب السنة الثالثة – قسم الإحصاء والبرمجة

2020


mohamed khatab

المقدمة:

لقد نشأ الإحصاء الرياضي من تقاطع الإحصاء والرياضيات، وهو وطيد الصلة بنظريتي الاحتمالات والعينات . ولذلك حرصنا عند إعداد هذه المحاضرات لتغطية مقرر (الاحصاء الرياضي), على أن تكون موضوعاتها مبسطة ومختصرة ومتفقة مع منهاج السنة الثالثة لطلاب الإحصاء في كلية الاقتصاد .

ولهذا فلقد عرضنا تلك الموضوعات ضمن خمسة فصول هي:

الفصل الأول: التقدير النقطي .

الفصل الثاني: التقدير المجالي .

الفصل الثالث: اختبارات الفرضيات البسيطة .

الفصل الرابع: تحليل التباين البسيط ز

الفصل الخامس: الانحدار الخطي البسيط .

ولقد استقينا هذه الفصول من المنشورات الجامعية التالية :

- العلي، إبراهيم محمد+ كابوس، أمل: الإحصاء الرياضي(1986) . من منشورات جامعة حلب.
- العلي,إبراهيم محمد+عكروش,محمد:الإحصاء التطبيقي(2005), من منشورات جامعة تشرين.
- العلي، إبراهيم محمد: أسس التحليل الإحصائي متعدد المتغيرات 2020، وهو من المنشورات الالكترونية للدكتور إبراهيم العلي على الموقع Dr-ALALI.com وغيره .

علماً بأن هذه المنشورات تعتمد على العديد من المراجع العالمية المذكورة فيها .

وإننا إذا نضع هذه المحاضرات بين أيدي طلابنا وزملائنا، لا ندعي إننا وصلنا بها إلى أي حد من الكمال والشمول، ولكننا نأمل أن نكون قد وفقنا في عرض موضوعاتها واستدراج أبحاثها واختيار تطبيقاتها. كما نأمل من القراء الكرام أن يوافونا بأية ملاحظة قد تكون مفيدة لنا عند إعادة طباعة أو نشر هذه المحاضرات .

المؤلفان

2020/5/25

ز

الفصل الأول : التقدير النقطي

1−1: تمهيد:

يتناول الاحصاء الرياضي قضايا تقدير معالم المجتمع parameters من خلال مؤشرات عينة عشوائية مسحوبة منه، فإذا كنا نريد تقدير متوسط (أو توقع) المجتمع μ المجهول فإننا نستخدم متوسط العينة $\bar{x}$ ، وإذا كنا نريد تقدير تباين المجتمع σ^2 المجهول فإننا نستخدم تباين العينة D^2 أو s^2 . ولكن هذه التقديرات قد تكون متحيزة أو قد تكون ذات خطأ كبير، وقد تكون غير متناسقة مع العينات الصغيرة ...الخ . لذلك كان لابد من وضع معايير لجودة التقدير قبل استنباط المقدرات التي سنستخدمها في تقدير معالم المجتمع .

ولتوضيح ذلك نفترض أننا نريد تقدير بعض خواص متحول X (كالمتوسط والتباين والنسبة) في مجتمع ما . فقمنا بسحب عينة عشوائية بحجم n عنصراً من عناصره وحصلنا منها على قياسات X التالية :

$$X:\ x_1\ ,x_2\ ,x_3\ ,\ldots.,x_i\ ,\ldots,x_n$$

ويجب أن نستخدم هذه القياسات بصيغة رياضية مناسبة للحصول على تقديرات خواص X المطلوبة .

فإذا كان هدفنا تقدير المتوسط العام لـ X في المجتمع والذي سنرمز له بـ μ , فإننا نجمع تلك القياسات ونقسمها على n (أي نأخذ المتوسط $\bar{x}$) فنحصل على تقدير μ المطلوب، ولكن هذه الصيغة ليست وحيدة، لأنه يمكننا تقدير المتوسط العام μ بأساليب وبصيغ أخرى، نذكر منها الوسيط والمنوال ومتوسط القيمتين الصغرى والعظمى ...الخ .

كما أنه يمكننا الحصول على تقديرات مختلفة من ذات الصيغة الرياضية بسحب عينات مختلفة من ذلك المجتمع (بحجم موحد n أو بحجوم مختلفة) . وذلك لأن عدد العينات التي يمكن سحبها عشوائياً وبالحجم n نفسه يساوي في حالة السحب بدون إعادة C_N^n عينة . ويساوي في حالة السحب مع الإعادة C_{N+n-1}^n عينة، حيث N عدد عناصر المجتمع و C_N^n عدد التوافيق بحجم n من N . وهكذا نجد أن القياسات التي سنحصل عليها من كل عينة تختلف عن قياسات العينات الأخرى . وبالتالي فإننا سنحصل على تقديرات مختلفة لمعالم المجتمع، حتى لو استخدمنا صيغة موحدة لحسابها، وذلك بسبب اختلاف العينات الممكنة نفسها .

ومما سبق نلاحظ أن عملية تقدير أي مؤشر إحصائي يمكن أن تعطينا قيماً متعددة ناتجة عن اختلاف العينات الممكنة، أو عن الصيغة الرياضية المستخدمة في عملية التقدير .

ومما سبق نستنتج أن تقدير أي مؤشر إحصائي في المجتمع، مثل θ , هو عبارة عن متحول عشوائي ويخضع لتوزيع احتمالي معين، وإن θ يكون داخلاً في تعريف ذلك التوزيع الاحتمالي.

1-2: تعاريف :

- **التقدير النقطي:** إن التقدير النقطي لأي مؤشر إحصائي θ في المجتمع للمتحول المدروس X , هو عبارة عن أية قيمة عددية من مجموعة القيم الممكنة, التي يمكن أن يعد كل منها بديلاً عن ذلك المؤشر θ . وتسنتج هذه القيم العددية من العينات الممكنة ووفق الصيغ الرياضية المختلفة .

- **المقدر** : هو الصيغة الرياضية التي تطبق على قياسات العينة $(x_1, x_2, x_3, \dots, x_n)$ للحصول على تقدير معين للمؤشر المجهول θ . وإذا رمزنا لتقدير θ بـ $\tilde{\theta}$. فإن الرمز $\tilde{\theta}$ سيكون تابعاً لقياسات العينة وفق الصيغة الرياضية المستخدمة ونكتبها على الشكل التالي :

$$\tilde{\theta} = H(x_1, x_2, x_3, \dots, x_n) \qquad (1-1)$$

حيث أن H هي صيغة المقدر المستخدم و $\tilde{\theta}$ هو التقدير الناتج .

وكمثال على ذلك نأخذ التقديرين التاليين لمتوسط وتباين X في المجتمع :

$$\tilde{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (2-1)$$
$$\tilde{\sigma}^2 = D^2 = \frac{1}{n}\sum (x_i - \bar{x})^2$$

- **التوزيع المشترك لقياسات العينة** : عندما نسحب عينة عشوائية بحجم n من المجتمع . فإنها تكون إحدى العينات الممكنة . وبما أننا لا نستطيع أن نحدد مسبقاً أي من هذه العينات أصبحت بين أيدينا , فإننا سنناقش الأمر بشكله العام . لذلك سنفترض أننا نتعامل مع جميع العينات الممكنة ذات الحجم n , والتي ستعطينا القياسات التالية :

i: 1 2 3 i n

X: x_{11} x_{12} x_{13} .... x_{1i} .... x_{1n} (قياسات العينة الأولى)

X: x_{21} x_{22} x_{23} .... x_{2i} .... x_{2n} (قياسات العينة الثانية)

X: x_{31} x_{32} x_{33} .... x_{3i} .... x_{3n} (قياسات العينة الثالثة)

... ...

... ...

وبذلك يمكننا أن ننظر إلى قياسات العمود الأول (قيم العنصر الأول في جميع العينات الممكنة)، على أنها قيم لمتحول عشوائي جديد نرمز له بـ X_1 ولنفترض أنه يخضع لتوزيع احتمالي $f_1(X_1)$. كما ننظر إلى قياسات العمود الثاني على أنها قيم لمتحول عشوائي آخر X_2 ويخضع لتوزيع $f_2(X_2)$ ، وإلى قياسات العمود الثالث على أنها متحول عشوائي ثالث X_3 ويخضع لتوزيع $f_3(X_3)$ ...الخ . وبالتالي نحصل على جملة من المتحولات العشوائية العشوائية المستقلة ونكتبها مع توزيعاتها الاحتمالية كما يلي:

$$\begin{matrix} X_1 & X_2 & X_3 & \dots & X_i & \dots & X_n \\ f_1(X_1) & f_2(X_2) & f_3(X_3) & \dots & f_i(X_i) & \dots & f_n(X_n) \end{matrix} \qquad (3-1)$$

وبما ان هذه المتحولات مستقلة فإن توزيعها المشترك يعطى بعلاقة الجداء التالية:

$$L(X_1\ \ X_2\ X_3\ \ldots\ X_n) = f_1(X_1) * f_2(X_2) * f_3(X_3) \ldots f_n(X_n) \qquad (4-1)$$

ولكن بما أن قيم كل من هذه المتحولات هي في الأصل قيم للمتحول المدروس X في نفس المجتمع، والذي يخضع لتوزيع معين هو $f(X)$ ، فإن كل من التوزيعات $f_i(X_i)$ سيكون لها نفس شكل التوزيع $f(X)$. وسيكون لها نفس القيم المميزة لـ X ، ولهذا فإنه يمكننا أن نعتبر التوزيع الأساسي $f(X)$ توزيعاً موحداً لجميع هذه المتحولات، ونكتفي باستخدام الرمز $f(X_i)$ عوضاً عن $f_i(X_i)$.

وعندما نريد أن نقدر أي مؤشر احصائي θ , فإن θ يكون داخلاً في الصيغة الرياضية للتوزيع $f(X)$ أو مرتبطاً بوسطائه . ولذلك سنرمز لتوزيع X في المجتمع بشكل عام بالرمز $f(X,\theta)$. وبذلك يكون θ داخلاً في الصيغة الرياضية لكل واحد من التوزيعات $f(X_i)$ للمتحولات المنفردة X_i ونكتبها كما يلي: $f(X_i,\theta)$.

وعندها فإن التوزيع المشترك للمتحولات السابقة $L(X_1\ \ X_2\ X_3\ \ldots\ X_n)$ يأخذ الشكل والرمز التالي:

$$L(X_1\ \ X_2\ X_3\ \ldots\ X_n\ \theta) = f(X_1,\theta) * f(X_2,\theta) \ldots \ldots f(X_n,\theta) = \prod_{i=1}^{n} f(X_i,\theta) \qquad (5-1)$$

وبما ان التقدير $\tilde{\theta}$ متحول عشوائي وتابع للقياسات $(x_1 , x_2 , x_3 , \ldots , x_n)$ حسب العلاقة (1-1) , فإن قانون توزيعه الاحتمالي سيكون مرتبطاً بالتوزيع المشترك $L(X_1\ \ X_2\ X_3\ \ldots\ X_n\ \theta)$ ، ومن ثم بالتوزيعات المنفردة $f(X_1,\theta)$ و $f(X_2,\theta)$ ...الخ . ولهذا فإننا سنستخدم التوزيع المشترك $L(X_1\ \ X_2\ X_3\ \ldots\ X_n\ \theta)$ في استنباط المقدرات (الصيغ الرياضية) التي سنطبقها على قياسات العينة للحصول على التقديرات اللازمة . ولكن قبل أن نتعرض لهذه القضايا سنقوم باستعراض بعض المعايير التي يجب أن تتوفر في التقدير $\tilde{\theta}$.

1-3: معايير جودة التقدير:

إن تقديراً $\tilde{\theta}$ للمؤشر θ . قد يكون قريباً جداً من θ ، وقد يكون بعيداً عنه . ولذلك تم وضع عدد من المعايير لضمان جودة التقدير $\tilde{\theta}$. وهذه المعايير هي:

1-3-1: عدم التحيز (Unbaised)

- **تعريف:** نقول عن التقدير $\tilde{\theta}$ إنه تقدير غير متحيز لـ $\tilde{\theta}$ ، إذا كان توقعه على جميع العينات الممكنة يساوي θ . أي إذا كان :

$$E(\tilde{\theta}) = \theta \qquad (6-1)$$

أما إذا كان : $E(\tilde{\theta}) \neq \theta$ وكان : $E(\tilde{\theta}) = \theta + \delta$

فإننا نقول عن $\tilde{\theta}$ إنه تقدير متحيز، وإن مقدار التحيز يساوي δ .

- **مثال** : لنفترض أنه لدينا متحولاً عشوائياً X . توقعه: $E(X) = \mu$ ، وتباينه: $Var\ (X) = \sigma^2$ ، وإننا نريد تقدير μ وσ^2 بواسطة متوسط العينة وتباينها $\bar{x}$ و D^2 على الترتيب. لذلك سحبنا عينة عشوائية بحجم n من عناصر المجتمع , ثم أخذنا قياسات X منها فحصلنا على القياسات التالية :

$$X: (x_1\ \ \ x_2\ \ \ x_{13}\ \ \ \ldots .\ x_i\ \ldots .\ x_n)$$

وحسبنا متوسط هذه القياسات (متوسط العينة) فكان يساوي:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (7-1)$$

ولدراسة فيما إذا كان $\bar{x}$ يمثل تقديراً غير متحيز للتوقع μ نأخذ توقعه على جميع العينات فنجد أن:

$$E(\bar{x}) = E\left[\frac{1}{n}\sum x_i\right] = \frac{1}{n}\sum E(x_i) = \frac{1}{n} * n\mu = \mu \qquad (8-1)$$

وهذا يعني أن متوسط العينة $\bar{x}$ هو تقدير غير متحيز للتوقع μ (المتوسط العام في المجتمع) .

ولدراسة فيما إذا كان تباين العينة المعرف بالعلاقة :

$$D^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 \qquad (9-1)$$

يمثل تقديراً غير متحيز لتباين المجتمع σ^2 نأخذ توقعه على جميع العينات الممكنة فنجد أن:

$$E(D^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2 = \frac{1}{n} * E\left\{\sum_{i=1}^{n}[(x_i - \mu) - (\bar{x} - \mu)]^2\right.$$

وبعد المعالجة نجد أن:

$$E(D^2) = \frac{1}{n}E\left[\sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2\right]$$

$$= \frac{1}{n}\left[\sum E(x_i - \mu)^2 - nE(\bar{x} - \mu)^2\right]$$

$$E(D^2) = \frac{1}{n}[n\sigma^2 - n\sigma_{\bar{x}}^2] = \sigma^2 - \sigma_{\bar{x}}^2 \qquad (10-1)$$

حيث رمزنا للتوقع الأخير بـ $\sigma_{\bar{x}}^2 = E(\bar{x} - \mu)^2$ للدلالة على تباين التقدير $\bar{x}$ عن μ . وهو يساوي حسب نظرية الاحتمالات $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ (في حالة السحب مع الإعادة)، وبالتعويض يكون لدينا:

$$E(D^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \neq \sigma^2 \qquad (11-1)$$

أي أن تباين العينة D^2 هو تقدير متحيز لتباين المجتمع σ^2 وإن مقدار التحيز يساوي $\delta = -\frac{\sigma^2}{n}$

وللتخلص من هذا التحيز نقوم بضرب التباين D^2 بالكسر $\frac{n}{n-1}$ فنحصل على تباين جديد للعينة يسمى بالتباين المصحح للعينة ونرمز له بالرمز S^2 ونكتب ذلك كما يلي:

$$S^2 = \frac{n}{n-1}D^2 = \frac{n}{n-1} * \frac{1}{n}\sum (x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (12-1)$$

وبسهولة يمكن البرهان على أن S^2 هو تقدير غير متحيز لـ σ^2 لأن :

$$E(S^2) = E\left[\frac{n}{n-1}D^2\right] = \frac{n}{n-1}E(D^2) = \frac{n}{n-1} * \frac{n-1}{n}\sigma^2 = \sigma^2$$

$$E(S^2) = \sigma^2 \qquad (13-1)$$

هذا يعني أن S^2 هو تقدير غير متحيز لـ σ^2 .

1-3-2: الاتساق (التماسك) (Consistent)

- **تعريف:** نقول عن التقدير $\tilde{\theta}$ إنه تقدير متسق لـ θ ، إذا كان $\tilde{\theta}$ ينتهي احتمالياً إلى θ عندما n تنتهي أو تقترب من اللانهاية، أي أنه يكون $\tilde{\theta}$ تقديراً متسقاً لـ θ ، إذا وجد مقابل أي عددين موجبين ε و∝ قيمة لـ n مثل n_0 بحيث يكون :

$$P\left(\left|\tilde{\theta}-\theta\right|<\varepsilon\right)\geq 1-\propto \qquad (14-1)$$

وذلك من أجل جميع n التي تكون $n\geq n_0$

وبتعبير آخر يكون التقدير $\tilde{\theta}$ تقديراً متسقاً لـ θ عندما يتحقق الشرط التالي:

$$P\left[\ \tilde{\theta}\xrightarrow[n\to N(\infty)]{}\theta\ \right]=1 \qquad (15-1)$$

- **مثال:إن متوسط العينة $\bar{x}$ هو تقدير متسق لمتوسط المجتمع μ لأن صيغة $\bar{x}$ تنتهي إلى صيغة μ كما يلي:**

$$P\left[\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i\xrightarrow[n\to N]{}\frac{1}{N}\sum_{i=1}^{N}x_i=\mu\right]=1 \qquad (16-1)$$

- **نظرية:** إذا كان $\tilde{\theta}$ تقديراً غير متحيز لـ θ وكان تباينه ينتهي إلى الصفر: $\sigma_{\tilde{\theta}}^2=E\left(\tilde{\theta}-\theta\right)^2\longrightarrow 0$. فإن $\tilde{\theta}$ يكون تقدير متسقاً لـ θ (حسب متراجحة تشيبشيف) .

1-3-3: الفعالية :

- **تعريف:** نقول عن التقدير $\tilde{\theta}$ إنه تقدير فعال لـ θ، إذا كان تباينه أصغر من تباينات جميع التقديرات الأخرى لـ θ . أي أن $\tilde{\theta}$ يكون فعالاً إذا كان :

$$\sigma_{\tilde{\theta}}^2\leq\sigma_{\tilde{\theta}_i}^2 \qquad \forall_i \qquad (17-1)$$

ويعد هذا المعيار مقياساً لتفضيل التقديرات المختلفة للمؤشر θ .

- **مثال: تقدير (ماركوف) لمتوسط المجتمع μ :**

لنفترض أننا سحبنا عينة عشوائية بحجم n من مجتمع ما وحصلنا منها على القياسات المستقلة التالية:

$$X: x_1\ \ x_2\ \ x_3\ \ldots\ x_i\ \ldots\ x_n$$

ونريد إيجاد تقدير غير متحيز وفعال لمتوسط المجتمع μ .

الحل: لإيجاد التقدير غير المتحيز والفعال $\tilde{\mu}$ للمتوسط μ يجب أن يتحقق لدينا الشرطان التاليان:

$$E(\tilde{\mu})=\mu \qquad (18-1)$$

$$\sigma_{\tilde{\mu}}^2=E(\tilde{\mu}-\mu)^2=Minimum \qquad (19-1)$$

ولذلك نفترض أن $\tilde{\mu}$ يرتبط مع القياسات $(x_1\ \ x_2\ \ x_3\ \ldots\ x_n)$ بواسطة علاقة خطية (أسهل العلاقات) مع الشكل التالي:

$$\tilde{\mu}=a_1x_1+a_2x_2+a_3x_3+a_nx_n=\sum_{i=1}^{n}a_ix_i \qquad (20-1)$$

حيث أن a_i هي مقادير عددية يجب تحديدها بحيث يتحقق لدينا الشرطان السابقان، ومن الشرط الأول نجد أن:

$$E(\tilde{\mu}) = E\left[\sum_{i=1}^{n} a_i x_i\right] = \sum_{i=1}^{n} a_i * E(x_i) = \sum_{i=1}^{n} a_i \mu = \mu \sum_{i=1}^{n} a_i$$

وهذا يعني أنه يجب أن يكون :

$$\mu \sum_{i=1}^{n} a_i = \mu$$

ومنها نستنتج أن المقادير a_i يجب أن تحقق العلاقة التالية:

$$\sum_{i=1}^{n} a_i = 1 \qquad (21-1)$$

ومن الشرط الثاني نجد أن :

$$\sigma_{\tilde{\mu}}^2 = var\left[\sum_{i=1}^{n} a_i x_i\right] = \sum_{i=1}^{n} a_i^2 * var(x_i) \qquad (22-1)$$

وبما أن $var(x_i) = \sigma^2$ وهو تباين المجتمع ويعد مقداراً ثابتاً فإننا نجد أن:

$$\sigma_{\tilde{\mu}}^2 = \sigma^2 \sum a_i^2 \Rightarrow \min \qquad (23-1)$$

ولجعل التباين $\sigma_{\tilde{\mu}}^2$ أصغر ما يمكن مع وجود الشرط $[\sum a_i = 1]$ نقوم بتشكيل تابع (لاغرانج) منهما كمايلي:

$$L = \sigma^2 \sum a_i^2 - \lambda\left[\sum a_i - 1\right] \qquad (24-1)$$

ثم نشتق بالنسبة للمقادير a_i ونضعها مساوية للصفر فنحصل على أن :

$$\frac{\partial L}{\partial a_i} = \sigma^2(2a_i) - \lambda[1-0] = 0$$

ومنها نحصل على أن:

$$\lambda = 2\sigma^2 a_i \qquad (25-1)$$

ثم نأخذ المجموع النوني للطرفين فنجد بعد الاستفادة من الشرط $[\sum a_i = 1]$ فنجد أن:

$$n\lambda = 2\sigma^2 \sum a_i = 2\sigma^2$$

أي أن:

$$\lambda = \frac{2\sigma^2}{n} \qquad (26-1)$$

وبالتعويض في العلاقة (1−25) نحصل على أن:

$$\frac{2\sigma^2}{n} = 2\sigma^2 * a_i \qquad (27-1)$$

ومنها نحصل على أن:

$$a_i = \frac{1}{n} \qquad : i \leq n$$

وبالتعويض في العلاقة الخطية (1−20) نحصل على أن التقدير المطلوب غير المتحيز والفعال لـ μ. هو :

$$\tilde{\mu} = \frac{1}{n}\sum x_i = \bar{x} \qquad (28-1)$$

وهو المتوسط الحسابي للعينة $\bar{x}$ والذي تباينه يساوي:

$$\sigma_{\tilde{\mu}}^2 = \sigma^2 \sum \left(\frac{1}{n}\right)^2 = \frac{\sigma^2}{n} \qquad (29-1)$$

وبما أن $\sigma_{\tilde{\mu}}^2 \longrightarrow 0$ عندما $n \longrightarrow 0$ فإن التقدير $\tilde{\mu} = \bar{x} = \frac{1}{n}\sum x_i$ هو تقدير متسق لـ μ. .

وبذلك يكون متوسط العينة $\bar{x}$ هو أفضل تقدير لمتوسط المجتمع μ .

ويمكننا اتباع الأسلوب نفسه للبرهان على أن التقدير غير المتحيز والفعال للتباين σ^2 هو تباين العينة المصحح المعرف سابقاً بالعلاقة :

$$\tilde{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

1−3−4: الكفاية :

- **تعريف:** نقول عن تقدير من الشكل: $\tilde{\theta} = \tilde{\theta}(x_1 x_2 \dots x_n)$ إنه تقدير كافٍ للمؤشر θ، إذا كان التوزيع الشرطي للحصول على أية عينة − بعد حساب $\tilde{\theta}$ − مستقلاً عن θ الحقيقية . وهذا يعني أنه عندما نضع الصيغة $\tilde{\theta}(x_1 x_2 \dots x_n)$ كتقدير لـ θ فإن هذه الصيغة تعني أن $\tilde{\theta}$ تمثل تكثيفاً لجميع قياسات العينة المسحوبة، أي أن هذا التقدير قد استخدم جميع المعلومات التي توفرها العينة . وبالعكس نجد أن التقدير :

$$\tilde{\theta} = \frac{x_{min} + x_{max}}{2} \qquad (30-1)$$

لا يستخدم جميع قياسات العينة . لذلك لا يعتبر تقديراً كافياً لمتوسط المجتمع μ . لأنه لم يستخدم غير قيمتين من تلك القياسات وهما: القيمة الصغرى والقيمة الكبرى .

1−4: الحد الأدنى لتباين التقدير (متراجحة كرامر− راو)− وتطبيقاته :

1−4−1: تمهيد رياضي :

نعلم أن معايير جودة التقدير $\tilde{\theta}$ لـ θ هي:

1− عدم التحيز: وهو أن يكون $E(\tilde{\theta}) = 0$ على جميع العينات الممكنة .

2−التماسك : وهو أن يكون $\tilde{\theta} \underset{n\to N}{\longrightarrow} \theta$ مثل: $\bar{x} \longrightarrow \bar{y}$: وهو يخص الصيغة الرياضية.

3−الفعالية: وهو أن يكون تباين $\tilde{\theta}$ أصغر ما يمكن، أي أن يكون $E(\tilde{\theta} - \theta)^2 \Longrightarrow Min$.

4− الكفاية: هو شرط استخدام جميع بيانات العينة .

إن المعيار الثالث يشترط أن يكون التباين $\sigma_{\tilde{\theta}}^2$ أو الخطأ المعياري $\sigma_{\tilde{\theta}}$ أصغر ما يمكن، وهذا يجعلنا نطرح على أنفسنا السؤال التالي: هل هناك حد أدنى لهذا التباين مثل σ_0^2 لا يمكن أن يكون أقل منه ؟ أي $\sigma_{\tilde{\theta}}^2 \geq \sigma_0^2$.

ولإيجاد هذا الحد الأدنى سنعالج الأمر كما يلي:

لنفترض أن X يخضع للتوزيع الاحتمالي $f(X,\theta)$ حيث θ هو وسيط مجهول في التوزيع ويجب تقديره من بيانات العينة . مثل متوسط المجتمع μ أو $\bar{y}$ أو تباينه σ^2 أو انحرافه σ ...الخ .

لذلك نسحب عينة من المجتمع بحجم n ونسجل قيم X فيها فنحصل على :

$$X:\ x_1\ x_2\ x_3\ x_4 \dots x_i \dots x_n$$

وإذا سحبنا جميع العينات الممكنة فإننا سنحصل على المتحولات المستقلة التالية:

$$X:\ X_1\ X_2\ X_3\ X_4 \dots X_i \dots X_n$$

وإن التوزيع المنفرد لكل منها هو $f(X_i\ \theta)$. أما التوزيع المشترك لها فيساوي :

$$L(X_1\ X_2\,X_3\ \dots\ X_n\,,\theta) = f(X_1\,,\theta) * f(X_2\,,\theta) * f(X_3\,,\theta) \dots f(X_n\,,\theta)$$

وسنرمز لهذا التوزيع المشترك اختصاراً بالرمز $L(X\,,\theta)$ وهو يتضمن الوسيط θ . وعلينا تقدير θ منه بحيث يكون تباين التقدير $\sigma^2_{\hat{\theta}}$ أصغر ما يمكن .

إن التوزيع المشترك $L(X\,,\theta)$ هو توزيع احتمالي معرف على منطقة D في الفضاء R^n ، وإنه لابد أن يحقق الشرط التالي :

$$\int_D L(X_1\ X_2 \dots X_n\,,\theta) dX_1 dX_2 \dots dX_n = \int_D L(X\,,\theta)\, dX = 1 \qquad (31-1)$$

وبفرض أن $L(X\,,\theta)$ قابل للاشتقاق بالنسبة لـ θ نقوم باشتقاق طرفي العلاقة (1−31) فنحصل على أن:

$$\int_D \frac{\partial L(X\,,\theta)}{\partial\theta} dx_1 dx_2 \dots dx_n = 0 \qquad (32-1)$$

ونظراً لصعوبة التعامل مع المشتق $\frac{\partial L(X\,,\theta)}{\partial\theta}$ نستبدله بالمشتق $\frac{\partial \ln L(X\,,\theta)}{\partial\theta}$ وذلك من خلال علاقته بمشتق اللوغاريتم وهي :

$$\frac{\partial \ln L}{\partial\theta} = \frac{L'}{L} = \frac{\frac{\partial L}{\partial\theta}}{L}$$

وهذا يعني أن :

$$\frac{\partial L}{\partial\theta} = \frac{\partial \ln L}{\partial\theta} * L \qquad (33-1)$$

نعوض (1−33) في (1−32) فنحصل على أن :

$$\int_D \frac{\partial \ln L(X\,,\theta)}{\partial\theta} * L *\ dx_1 dx_2 \dots dx_n = 0 \qquad (34-1)$$

وهذا يعني أن التوقع الرياضي للمشتق $\left(\frac{\partial \ln L}{\partial\theta}\right)$ يساوي الصفر, أي ان:

$$E\left(\frac{\partial \ln L}{\partial\theta}\right) = 0 \qquad (35-1)$$

وإن هذا التوقع مأخوذ على جميع لعينات الممكنة بحجم n .

ثم نقوم باشتقاق العلاقة (1−34) مرة أخرى بالنسبة لـ θ فنحصل على ما يلي :

$$\int_D \left(\frac{\partial^2 \ln L}{\partial^2\theta} L + \frac{\partial L}{\partial\theta} * \frac{\partial \ln L}{\partial\theta}\right) dX = 0$$

وبالاستفادة من (1−33) نحصل على :

$$\int_D \left(\frac{\partial^2 \ln L}{\partial^2\theta} * L + \frac{\partial \ln L}{\partial\theta} * L * \frac{\partial \ln L}{\partial\theta}\right) dX = 0$$

$$\int_D \frac{\partial^2 \ln L}{\partial^2\theta} * L * dX = -\int \left(\frac{\partial \ln L}{\partial\theta}\right)^2 * L * dX$$

وبما أن: $E\left(\frac{\ln L}{\partial\theta}\right) = 0$ حسب العلاقة (1−35) نكتب الطرف الأيمن كما يلي:

$$\int_D \frac{\partial^2 \ln L}{\partial^2\theta} * L * dX = -\int_D \left[\frac{\partial \ln L}{\partial\theta} - 0\right]^2 * L * dX$$

أي أن:

$$-E\left[\frac{\partial^2 \ln L}{\partial^2\theta}\right] = +E\left[\frac{\partial \ln L}{\partial\theta} - 0\right]^2 = var\left(\frac{\partial \ln L}{\partial\theta}\right)$$

ومنها نستنتج أن:

$$\boxed{-E\left[\frac{\partial^2 \ln L}{\partial^2\theta}\right] = E\left(\frac{\partial \ln L}{\partial\theta}\right)^2 = var\left(\frac{\partial \ln L}{\partial\theta}\right)} \qquad \mathbf{(36-1)}$$

ونعود الآن بعد هذا التمهيد إلى إيجاد الحد الأدنى لتباين أي تقدير $\tilde{\theta}$ لـ θ.

1−4−2: إيجاد الحد الأدنى لتباين التقدير $\tilde{\theta}$ (متراجحة كرامر− راو) :

لنفترض الآن أن t هو تقدير غير متحيز لـ θ , أو بصورة عامة نفترض أن t هو تقدير غير متحيز لتركيب من θ هو التابع $\tau(\theta)$, فعندها يكون لدينا من شرط عدم التحيز ما يلي:

$$E(t) = \tau(\theta)$$

وهذا يعني أن:

$$\int_D t * L * dX = \tau(\theta)$$

نقوم باشتقاق الطرفين بالنسبة لـ θ فنجد أن:

$$\int_D t \frac{\partial L}{\partial\theta} * dX = \tau'(\theta)$$

وبالاستفادة من العلاقة (1−33) نحصل على أن:

$$\int t * \frac{\partial \ln L}{\partial\theta} * L * dX = \tau'(\theta) \qquad \mathbf{(37-1)}$$

وبضرب طرفي العلاقة (1−34) بالتركيب $\tau(\theta)$ نحصل على أن:

$$\int \tau(\theta) * \frac{\partial \ln L}{\partial\theta} * L * dX = 0 \qquad \mathbf{(38-1)}$$

وبطرح العلاقة (1−38) من (1−37) مع ملاحظة (1−35) نحصل على أن:

$$\int_D [t - \tau(\theta)] * \left[\frac{\partial \ln L}{\partial\theta} - 0\right] * L * dX = \tau'(\theta) \qquad \mathbf{(39-1)}$$

وبما أن الطرف الأيسر هو التباين المشترك COV للمتحولين المذكورين فيه نستخلص منها أن :

$$COV\left(t,\frac{\partial \ln L}{\partial \theta}\right)=\tau'(\theta) \qquad \mathbf{(40-1)}$$

ونحن نعلم من متراجحة (كوشي– شوارز) لمعامل الارتباط r بين أي متحولين X وY أن:

$$r=\frac{COV(X,Y)}{\sigma_x,\sigma_y}\leq 1$$

وبالتربيع نجد أن:

$$r^2=\frac{COV^2(X,Y)}{\sigma_x^2,\sigma_y^2}\leq 1$$

أي أن:

$$COV^2(X,Y)\leq \sigma_x^2,\sigma_y^2 \qquad (41-1)$$

وبتطبيق ذلك على الطرف الأيسر من العلاقة (1–40) نجد أن :

$$COV^2\left(t,\frac{\partial \ln L}{\partial \theta}\right)\leq \sigma_t^2 * Var\left(\frac{\partial \ln L}{\partial \theta}\right) \qquad (41\,a-1)$$

ومن (1–40) نجد أن:

$$\tau'^2(\theta)\leq \sigma_t^2 * Var\left(\frac{\partial \ln L}{\partial \theta}\right)$$

ومنها نجد أن تباين التقدير t وهو σ_t^2 يحقق المتراجحة التالية:

$$\sigma_t^2\geq \frac{\tau'^2(\theta)}{Var\left(\frac{\partial \ln L}{\partial \theta}\right)} \qquad (42-1)$$

ومن (1–36) نجد أيضاً أن:

$$\boxed{\sigma_t^2\geq \frac{\tau'^2(\theta)}{-E\left[\frac{\partial^2 \ln L}{\partial^2 \theta}\right]}\Bigg|=\sigma_0^2} \qquad (43-1)$$

وهنا نلاحظ أن الطرف الأيمن هو الحد الأدنى للتباين σ_t^2 للتقدير t وسنرمز له بالرمز σ_0^2 فنحصل على أن $\sigma_t^2\geq \sigma_0^2$.

إن العلاقة (1–43) تسمى متراجحة (كرامر– راو) وهي متراجحة مهمة جداً في معالجة مسائل التقدير .

- **حالة خاصة:** إذا كان التركيب المستخدم $\tau(\theta)$ مساوياً لـ θ , أي كان $\tau(\theta)=\theta$, فإنه يكون لدينا $\tau'^2(\theta)=1$ وبالتعويض في العلاقة (1–43) نحصل على المتراجحة التالية:

$$\boxed{\sigma_t^2\geq \frac{1}{-E\left[\frac{\partial^2 \ln L}{\partial^2 \theta}\right]}} \qquad (44-1)$$

ملاحظة: يمكن صياغة العلاقتين (1–43) و(1–44) بدلالة التوزيع المنفرد لـ X وهو $f(X,\theta)$، وذلك كما يلي:

$$\ln[L(X_1\ X_2\,X_3\ \dots\ X_n\,,\theta)] = \ln[f(X_1\,,\theta) * f(X_2\,,\theta) * f(X_3\,,\theta) \dots f(X_n\,,\theta)]$$

$$\ln L = \ln\prod f(X_i\,,\theta) = \sum_{i=1}^{n} \ln f(X_i\,,\theta) \qquad (45-1)$$

وعندما نأخذ المشتق الثاني للطرفين في (1−45) نحصل على أن:

$$\frac{\partial^2 \ln L}{\partial^2\theta} = \sum_{i=1}^{n}\frac{\partial^2 \ln f(X_i\,,\theta)}{\partial^2\theta} = n * \frac{\partial^2 \ln f(X\,,\theta)}{\partial^2\theta}$$

وذلك لأن التوزيعات $f(X_i\,,\theta)$ متساوية وتساوي $f(X\,,\theta)$.

وبالتعويض في العلاقة (1−43) نحصل على المتراجحة التالية:

$$\boxed{\sigma_t^2 \geq \frac{\tau'^2(\theta)}{-nE\left[\frac{\partial^2 \ln f(X\,,\theta)}{\partial^2\theta}\right]}} \qquad \mathbf{(46-1)}$$

وعند معالجة الحالة الخاصة التي يكون فيها $\tau(\theta) = \theta$. نحصل على الشكل التالي:

$$\boxed{\sigma_t^2 \geq \frac{1}{-nE\left[\frac{\partial^2 \ln f(X\,,\theta)}{\partial^2\theta}\right]}} \qquad \mathbf{(47-1)}$$

1−4−3: خواص التقدير الذي يبلغ تباينه حده الأدنى $\boldsymbol{\sigma_0^2}$:

- تعريف: نقول عن كل تقدير غير متحيز t للمقدار $\tau(\theta)$ إنه تقدير ذو تباين يبلغ الحد الأدنى σ_0^2 إذا كان $\sigma_t^2 = \sigma_0^2$, أي إذا كان تباين التقدير t يساوي حده الأدنى :

$$\sigma_t^2 = \frac{\tau'^2(\theta)}{-E\left[\frac{\partial^2 \ln L}{\partial^2\theta}\right]} = \sigma_0^2 \qquad (48-1)$$

أو إذا كانت لدينا الحالة الخاصة : $t = \tau(\theta) = \theta$ وكان تباين t يساوي:

$$\sigma_t^2 = \frac{1}{-E\left[\frac{\partial^2 \ln L}{\partial^2\theta}\right]} \qquad (48a-1)$$

وعندها يكون ذلك التقدير فعالاً بشكل مطلق , وإن هذه الحالة لا تحدث إلا عندما تصبح المتراجحة (1−41) على شكل مساواة كما يلي:

$$COV\left(t\,,\frac{\partial \ln L}{\partial\theta}\right) = \sigma_t^2 * Var\left(\frac{\partial \ln L}{\partial\theta}\right)$$

$$E\left[\left(t-\tau(\theta)\right)\left(\frac{\partial \ln L}{\partial\theta} - 0\right)\right] = \sigma_t^2 * E\left(\frac{\partial \ln L}{\partial\theta} - 0\right)^2$$

وهذا يعني أن قيمة معامل الارتباط r بين المتحولين t و$\left(\frac{\partial \ln L}{\partial\theta}\right)$ تساوي الواحد، ومن هذا نستنتج أن المتحولين المذكورين t و$\left(\frac{\partial \ln L}{\partial\theta}\right)$ يرتبطان بعلاقة خطية تناسبية من الشكل :

$$\frac{\partial \ln L}{\partial\theta} = A[t-\tau(\theta)] \qquad \mathbf{(49-1)}$$

حيث أن: A هو عامل التناسب وهو غير متعلق ببيانات العينة $(x_1 \;\; x_2 \dots x_n)$, ولكنه يمكن أن يكون A تابعاً للوسيط θ . لذلك نكتب علاقة التناسب السابقة كما يلي :

$$\frac{\partial \ln L}{\partial \theta} = A(\theta)[t - \tau(\theta)] \qquad (50-1)$$

ولحساب مقام العلاقة (1−48) نحسب المشتق الثاني فنجد أن:

$$\frac{\partial^2 \ln L}{\partial^2 \theta} = A'(\theta)[t - \tau(\theta)] - \tau'(\theta) * A(\theta)$$

ثم نحسب التوقع الرياضي على جميع العينات للطرفين فنجد أن:

$$E\left(\frac{\partial^2 \ln L}{\partial^2 \theta}\right) = A'(\theta)E[t - \tau(\theta)] - E[\tau'(\theta) * A(\theta)]$$

وبما أن t تقدير غير متحيز لـ $\tau(\theta)$ وأن $\tau(\theta)$ و $A(\theta)$ غير تابعين لبيانات العينة فإن:

$$E[t - \tau(\theta)] = E(t) - \tau(\theta) = 0$$
$$E[\tau'(\theta) * A(\theta)] = \tau'(\theta) * A(\theta)$$

وهكذا نجد أن:

$$E\left(\frac{\partial^2 \ln L}{\partial^2 \theta}\right) = 0 - \tau'(\theta) * A(\theta) \qquad (51-1)$$

وبالتعويض في العلاقة (1−48) نحصل على أن σ_t^2 في هذه الحالة يساوي :

$$\sigma_t^2 = \frac{\tau'^2(\theta)}{\tau'(\theta) * A(\theta)} = \frac{\tau'(\theta)}{A(\theta)} \qquad (52-1)$$

وبما أن $\sigma_t^2 > 0$ فإننا نأخذ القيمة المطلقة للطرف الأيمن ونكتب العلاقة (1−52) على الشكل التالي :

$$\sigma_t^2 = \left|\frac{\tau'(\theta)}{A(\theta)}\right| \qquad (53-1)$$

ملاحظة: عندما تكون لدينا الحالة الخاصة $\tau(\theta) = \theta$ فإن التباين σ_t^2 الذي يبلغ حده الأدنى يساوي:

$$\sigma_t^2 = \frac{1}{|A(\theta)|} \quad \Longrightarrow \quad |A(\theta)| = \frac{1}{\sigma_t^2} \qquad (54-1)$$

أي أن تباين التقدير الذي يبلغ حده الأدنى يساوي مقلوب عامل التناسب $A(\theta)$ مأخوذاً بالقيمة المطلقة . ويستفاد من هذه العلاقة في حساب قيمة ذلك التباين مباشرة من العلاقة (1−50) بعد إيجادها من التوزيع .

1−4−4: التوزيعات التي تقبل تقديرات تباينها يدرك الحد الأدنى:

لقد اشرنا إلى أن التقدير غير المتحيز t للمقدار $\tau(\theta)$ يبلغ حده الأدنى عندما يكون المتحولان $\frac{\partial \ln L}{\partial \theta}$ و t متناسبين خطياً ، أي مرتبطين بعلاقة من الشكل :

$$\frac{\partial \ln L}{\partial \theta} = A(\theta)[t - \tau(\theta)] \qquad (55-1)$$

إن هذه العلاقة تعطينا إمكانية لتحديد نوعية التوزيعات الاحتمالية التي تقبل تقديرات بتباين يبلغ حده الأدنى لذلك نكتب العلاقة السابقة كما يلي :

$$\frac{\partial \ln L}{\partial \theta} = A(\theta) * t - A(\theta)\,\tau(\theta)$$

ثم نكامل الطرفين بالنسبة لـ θ فنحصل على أن:

$$\ln L = t * B(\theta) + H(\theta) + C(x_1 x_2 \dots x_n)$$

ومنه نجد أن شكل التوزيع المشترك $L(x_1 x_2 \dots x_n , \theta)$ الذي يقبل تقديرات لها تباينات تبلغ حدها الأدنى, يجب أن يكون لها شكل الصيغة الأسية التالية:

$$L = e^{t*B(\theta)+H(\theta)} * e^{C(x_1 x_2 \dots x_n)}$$
$$L = L_1 e^{t*B(\theta)+H(\theta)} \qquad (56-1)$$

حيث L_1 هو تابع لـ $(x_1 x_2 \dots x_n)$ فقط و$B(\theta)$ و$H(\theta)$ تابعين لـ θ فقط .

وهذا يعني أن التوزيعات ذات الصيغ الأسية هي التي تقبل تقديرات يبلغ تباينها الحد الأدنى، والتي تكون مؤلفة من حدين منفصلين أحدهما تابع لبيانات العينة $(x_1 x_2 \dots x_n)$ والآخر تابع للوسيط θ .

ولكن حتى يأخذ التوزيع المشترك $L(x_1 x_2 \dots x_n , \theta)$ هذه الصيغة يجب أن يكون للتوزيع المنفرد والموحد $f(X , \theta)$ نفس الشكل الأسي . أي يجب أن يكون له صيغة من الشكل:

$$f(X_i , \theta) = \ell_1 e^{t'*B(\theta)+h(\theta)} \qquad (57-1)$$

حيث أن:

$$L_1 = \ell_1^n \qquad t = \sum t' \qquad H(\theta) = n * h(\theta)$$

مثال 1−1: إذا كان σ معلوماً فأوجد التقدير غير المتحيز الذي يبلغ تباينه حده الأدنى للوسيط θ في التوزيع الطبيعي التالي:

$$f(X , \theta) = \frac{1}{\sigma * \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\theta}{\sigma}\right)^2} \qquad (58-1)$$

الحل: نسحب عينة بحجم n من المجتمع المذكور فنجد أن التوزيع المشترك لقياسات هذه العينة هو:

$$L(X_1 X_2 \dots X_n , \theta) = \prod_{i=1}^{n} f(X_i , \theta) = \frac{1}{\left(\sigma * \sqrt{2\pi}\right)^n} * e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\theta}{\sigma}\right)^2} \qquad (59-1)$$

وبذلك نجد أن $\ln L$ يساوي :

$$\ln L = n * \ln\left(\frac{1}{\sigma * \sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \theta}{\sigma}\right)^2 \qquad (60-1)$$

وباشتقاق الطرفين بالنسبة لـ θ نحصل على أن:

$$\frac{\partial \ln L}{\partial \theta} = 0 - \frac{1}{2}\sum_{i=1}^{n} 2\left(\frac{x_i - \theta}{\sigma}\right)\left(\frac{-1}{\sigma}\right)$$
$$\frac{\partial \ln L}{\partial \theta} = \frac{1}{\sigma^2}\left[\sum x_i - n\theta\right]$$

والتي يمكن كتابتها على الشكل التالي :

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\sigma^2}\left[\frac{\sum x_i}{n} - \theta\right] = \frac{n}{\sigma^2}[\bar{x} - \theta]$$
$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\sigma^2}[\bar{x} - \theta] = A(\theta)(t - \theta) \qquad \mathbf{(61-1)}$$

وبمقارنة هذه العلاقة مع العلاقة (1−55) نجد أنها تأخذ نفس الشكل ومنها نستخلص أن:

$$A(\theta) = \frac{n}{\sigma^2} \qquad t = \bar{x} \qquad \tau(\theta) = \theta \qquad (62-1)$$

وهذا يعني أن $\bar{x}$ هو التقدير غير المتحيز لـ θ وإن تباين التقدير $\bar{x}$ لـ θ يبلغ حده الأدنى , والذي يساوي في هذه الحالة:

$$\sigma_{\bar{x}}^2 = \frac{1}{|A(\theta)|} = \frac{\sigma^2}{n} \qquad (63-2)$$

كما نستنتج أن التوزيع الطبيعي يقبل تقديرات لـ θ تبلغ تبايناتها الحد الأدنى .

مثال 1−2: إذا كان X خاضعاً لتوزيع بواسون المتضمن الوسيط θ كما يلي :

$$f(X, \theta) = \frac{\theta^x}{x_i} e^{-\theta} \qquad (64-1)$$

والمطلوب إيجاد التقدير غير المتحيز لـ θ والذي يبلغ تباينه الحد الأدنى .

الحل: نشكل التوزيع المشترك، فنجد أن:

$$L(X, \theta) = \frac{\theta^{\sum x_i}}{\prod x_i} e^{-n\theta} \qquad (65-1)$$

$$\ln L = \left(\sum x_i\right) * \ln\theta - n\theta - \ln\left(\prod x_i\right)$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{\sum x_i}{\theta} - n = \frac{\sum x_i - n\theta}{\theta} = \frac{n}{\theta}(\bar{x} - \theta) \qquad (66-1)$$

وهذا يتوافق مع العلاقة (1−55), ومنها نستنتج مباشرة أن تقدير θ هو $t = \bar{x} = \tilde{\theta}$ وأن تباينه هو $\sigma_t^2 = \frac{1}{|A(\theta)|} = \frac{\theta}{n}$ ، وتقديره $\tilde{\sigma}_t^2 = \frac{\bar{x}}{n}$ وتباين التقدير $\tilde{\theta} = \bar{x}$ يبلغ حده الأدنى .

مثال 1−3: إذا كان X خاضعاً للتوزيع الأسي التالي:

$$f(X) = \lambda e^{-\lambda x} \quad : \quad 0 \le x < \infty \quad : \quad \lambda > 0 \qquad (67-1)$$

وأوجد التقدير غير المتحيز والفعال لـ λ.

الحل: نشكل التوزيع المشترك كما يلي:

$$L(X, \lambda) = \prod_{i=1}^{n} f(x_i) = \lambda^n * e^{-\lambda \sum x_i} \qquad (68-1)$$

ثم نأخذ لوغاريتم الطرفين فنحصل على أن:

$$\ln L(X, \lambda) = n \ln\lambda - \lambda * \sum x_i$$

ثم نشتق بالنسبة لـ λ فنجد أن:

$$\frac{\partial \ln L}{\partial \lambda} = n\left(\frac{1}{\lambda}\right) - \sum x_i = \frac{n}{\lambda} - n\bar{x} = \frac{n}{\lambda}(1 - \bar{x} * \lambda)$$

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n * \bar{x}}{\lambda}\left(\frac{1}{\bar{x}} - \lambda\right) \qquad (69-1)$$

وهي علاقة مشابهة للعلاقة

$$\frac{\partial \ln L}{\partial \lambda} = A(\theta)(t-\theta)$$

ومن المقارنة فنستخلص أن تقدير λ هو المقدار $t = \frac{1}{\bar{x}}$ أي أن: $\tilde{\lambda} = \frac{1}{\bar{x}}$

وإن تباين ذلك التقدير يبلغ حده الأدنى ويساوي:

$$\sigma_t^2 = \frac{\lambda}{n\bar{x}} = \frac{1}{n\bar{x}^2} \qquad (70-1)$$

مثال 1–4: لنفترض أن X يخضع لتوزيع غاما $\Gamma_P(x)$ التالي :

$$\Gamma_P(x) = \frac{\propto^P * x^{P-1} * e^{-\propto x}}{\Gamma(P)} \qquad (71-1)$$

والمطلوب إيجاد تقدير غير متحيز وفعال للوسيط ∝ .

الحل: نشكل التوزيع المشترك كما يلي:

$$L(X, \propto) = \left[\frac{1}{\Gamma(P)}\right]^n * \propto^{nP} * \left(\prod x_i\right)^{P-1} * e^{-\propto \sum x_i} \qquad (72-1)$$

ثم نأخذ لوغاريتم الطرفين فنحصل على أن:

$$\ln L(X, \propto) = n * \ln\left(\frac{1}{\Gamma(P)}\right) + nP \ln \propto + (P-1)\ln \prod x_i - \propto \sum x_i$$

ثم نشتق بالنسبة لـ ∝ . علماً بأن $\Gamma(P)$ هو قيمة عددية (قيمة تكامل $\Gamma(P)$) فنجد أن:

$$\frac{\partial \ln L}{\partial \propto} = 0 + \frac{nP}{\propto} + 0 - \sum x_i = \frac{nP}{\propto} - n\bar{x} = \frac{nP - n\bar{x} \propto}{\propto}$$

$$= \frac{n\bar{x}}{\propto}\left(\frac{P}{\bar{x}} - \propto\right) \qquad (73-1)$$

وهذا يتناسب مع العلاقة الخطية $\frac{\partial}{\partial \propto} \ln L = A(\theta)\big(t - \tau(\theta)\big)$

ومنها نستنتج أن $\tilde{\propto} = t = \frac{P}{\bar{x}}$ وأن تباينه هو $\sigma_{\tilde{\propto}}^2 = \frac{\propto}{n\bar{x}}$ ، وهو يبلغ حده الأدنى التالي:

$$\sigma_0^2 = \frac{\propto}{n\bar{x}} = \frac{P}{n\bar{x}^2} \qquad (74-1)$$

مثال 1–5: أوجد تقدير P في التوزيع الثنائي التالي (لتجربة واحدة) المعرف بالعلاقة :

$$f(x, P) = C_1^x P^x * (1-P)^{1-x} \qquad (75-1)$$

الحل: نشكل التوزيع المشترك فنجد أن:

$$L(X, P) = [C_1^x]^n P^{\sum x_i} * (1-P)^{n-\sum x_i} \qquad (76-1)$$

ثم نأخذ لوغاريتم الطرفين فنجد أن:

$$\ln L(X, P) = n \ln[C_1^x] + \left(\sum x_i\right) * \ln P + \left(n - \sum x_i\right) * \ln(1-P)$$

ثم نشتق بالنسبة لـ P فنحصل على:

$$\frac{\partial \ln L}{\partial P} = 0 + \sum x_i \left(\frac{1}{P}\right) + \left(n - \sum x_i\right) * \left(\frac{-1}{1-P}\right) \qquad (77-1)$$

$$= \frac{\sum x_i}{P} + \frac{(\sum x_i - n)}{1-P} = \frac{(1-P)\sum x_i + P(\sum x_i - n)}{P(1-P)}$$

$$= \frac{\sum x_i - P\sum x_i + P\sum x_i - n*P}{P(1-P)} = \frac{\sum x_i - nP}{P(1-P)}$$

$$= \frac{n}{P(1-P)}(\bar{x} - P) = A(\theta)\big(A - \tau(\theta)\big) \qquad (78-1)$$

وبالمقارنة من العلاقة المطلوبة (1-55) نستخلص أن:

التقدير غير المتحيز لـ P هو $\bar{x}$ أي $\tilde{P} = \bar{x}$ وتباينه يبلغ حده الأدنى وهو: $\sigma_{\bar{x}}^2 = \frac{P(1-P)}{n}$

1-5: طرائق التقدير الاحصائي لمعالم المجتمع:

لقد عالجنا في الفقرات السابقة الشروط التي يجب أن تتوفر في التقديرات $\tilde{\theta}$، ولم نتعرض إلى كيفية الحصول على تلك التقديرات من التوزيعات الاحتمالية المشتركة, وبحيث يتم تحقيق كل أو بعض معايير جودة التقدير. ولقد تم استنباط العديد من الطرائق الرياضية لاستخراج الصيغ الرياضية (المقدرات), التي يمكن استخدامها لحساب التقديرات $\tilde{\theta}$ من قياسات العينة $(x_1 x_2\ x_3 \dots x_n)$.وأهم هذه الطرائق هي:

- طريقة العزوم- طريقة الامكانية العظمى- طريقة (بايز) .
- طريقة المربعات الصغرى- طريقة χ^2 الصغرى .

1-5-1: طريقة العزوم :

وهي من أقدم الطرائق المستخدمة في الحصول على التقديرات المختلفة، وتعتمد هذه الطريقة على مبدأ المطابقة بين العزوم الابتدائية أو المركزية للمتحول X في كل من المجتمع والعينة المسحوبة منه، واستخراج مقدرات المؤشرات من معادلات المطابقة المذكورة .

- فإذا كان X خاضعاً للتوزيع الاحتمالي $f(x\,,\theta)$. وكان التوزيع $f(x\,,\theta)$ يتضمن مؤشراً واحداً مجهولاً هو θ . فإن العزم الابتدائي الأول لـ X حسب ذلك التوزيع يساوي :

$$\mu_1(x) = \int_{-\infty}^{+\infty} x * f(x\,,\theta)dx \qquad \mathbf{(79-1)}$$

ومن جهة أخرى يمكننا حساب هذا العزم الأول من قياسات العينة باستخدام العلاقة:

$$\mu_1'(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \qquad \mathbf{(80-1)}$$

وبإجراء المطابقة بين هذين العزمين نحصل منهماعلى المعادلة التالية:

$$\tilde{\mu}_1(x) = \mu_1'(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \qquad (81-1)$$

ثم نقوم بمعالجة هذه المعادلة فنحصل على صيغة رياضية لتقدير المؤشر θ بدلالة $\mu_1'(x)$ ، أي بدلالة قياسات العينة $(x_1 x_2 \dots x_n)$.

- **أما إذا كان التوزيع الاحتمالي لـ X يتضمن k مؤشراً مجهولاً مثل $\theta_1\theta_2 \dots \theta_k$، فعندئذ يكون له الشكل التالي:**

$$f(X \ \ \theta_1\theta_2 \dots \theta_k) \qquad (82-1)$$

ولإيجاد مقدرات للمؤشرات $(\theta_1\theta_2 \dots \theta_k)$ فإننا نحتاج إلى k معادلة مستقلة . لذلك نقوم بمطابقة العزم الابتدائي الأول والعزوم المركزية حتى المرتبة k في المجتمع والعينة، أي نجعل العزوم النظرية التالية :

$$\mu_1(x) = \int_{-\infty}^{+\infty} x f(x\,, \theta_1\theta_2 \dots \theta_k)dx \qquad (83-1)$$

$$M_r(x) = \int_{-\infty}^{+\infty} (x-\mu_1)^r f(x\,, \theta_1\theta_2 \dots \theta_k)dx \qquad (84-1)$$

(**حيث أن $r : 2\ 3 \dots k$**). ثم نحسب العزوم المقابلة لها في العينة والتي تحسب من العلاقات التالية:

$$\mu_1'(x) = \frac{1}{n}\sum x_i = \bar{x}$$
$$M_r'(x) = \frac{1}{n}\sum (x_i - \bar{x})^r \qquad (85-1)$$

(**حيث $r : 2\ 3 \dots k$**). وبإجراء المطابقات بين هذه العزوم نحصل على جملة المعادلات التالية:

$$\tilde{\mu}_1(x) = \mu_1'(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$
$$\widetilde{M}_r(x) = M_r'(x) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^r \qquad \mathbf{(86-1)}$$

(حيث أن: $r: 2\ 3\ 4 \dots k$). وهي جملة معادلات مؤلفة من k معادلة، يمكننا إيجاد حلها المشترك وإيجاد التقديرات المناسبة لكل من المؤشرات المجهولة $\theta_1\theta_2\ \theta_3 \dots \theta_k$ بدلالة قياسات العينة $(x_1x_2 \dots x_n)$.

مثال1−6: أوجد بطريقة العزوم تقديراً لكل من توقع وتباين متحول X يخضع للتوزيع الطبيعي العام $N(\mu\,, \sigma^2)$، وذلك من خلال قياسات العينة المسحوبة $(x_1x_2\ x_3 \dots x_n)$.

الحل: إن التوزيع الاحتمالي لـ X يكتب على الشكل التالي:

$$f(x\,, \mu\,, \sigma^2) = \frac{1}{\sqrt{2\pi} * \sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (87-1)$$

ولإيجاد تقدير لـ μ ولـ σ^2 بطريقة العزم نقوم بحساب العزم الابتدائي الأول لـ X فنجد أنه يساوي:

$$\mu_1(x) = \int_{-\infty}^{+\infty} x * f(x\,, \mu\,, \sigma^2)dx = \mu \qquad (88-1)$$

ثم نقوم بحساب العزم المركزي الثاني فنجد أنه يساوي:

$$M_2(x) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x\,, \mu\,, \sigma^2)dx = \sigma^2 \qquad (89-1)$$

ثم نقوم بحساب العزمين المقابلين لهما في العينة حسب العلاقة (1−85) فنحصل على أن:

$$\mu_1'(x) = \frac{1}{n}\sum x_i = \bar{x}$$
$$M_2'(x) = \frac{1}{n}\sum (x_i - \bar{x})^2 = D^2 \qquad (90-1)$$

وبإجراء المطابقة بين هذين العزمين المتقابلين نحصل على المعادلتين التاليتين:

$$\tilde{\mu} = \tilde{\mu}_1(x) = \mu_1'(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \qquad (91-1)$$
$$\widetilde{\sigma^2} = \widetilde{M}_2(x) = M_2'(x) = \frac{1}{n}\sum (x_i - \bar{x})^2 = D^2$$

ووبذلك نكون قد حصلنا من هاتين المعادلتين على صيغتين رياضيتين لتقدير المؤشرين (μ, σ^2) ، بدلالة قياسات العينة قياسات العينة $(x_1 x_2 \dots x_n)$ ،

1-5-2: طريقة الإمكانية العظمى:

وتعتمد هذه الطريقة على تقدير معالم المجتمع $(\theta_1 \theta_2 \dots \theta_k)$ من التوزيع الاحتمالي المشترك لقياسات العينة $(x_1 x_2 \dots x_n)$ في النقطة التي تقابل المنوال، أي في النقطة التي تقابل أكبر قيمة لذلك التوزيع (قيمة عظمى له) . ولهذا فإنها تعتمد على حساب المشتقات الأولى بالنسبة للمؤشرات المجهولة ووضعها مساوية للصفر .

فنحصل على k معادلة مستقلة، نقوم بحلها فنحصل على التقديرات اللازمة حسب الحالات التالية :

أ- الحالة التي يكون لدينا مؤشراً واحداً θ مجهولاً :

وعندها فإن التوزيع الاحتمالي لـ X يكتب كما يلي $f(x, \theta)$، أي أنه يتضمن مؤشراً واحداً θ فإن التوزيع المشترك لقياسات العينة $(x_1 x_2 \dots x_n)$ يعطى بواسطة الجداء التالي:

$$L(x_1 x_2 \dots x_n\ \theta) = f(x_1, \theta) * f(x_2, \theta) \dots f(x_n, \theta)$$

وبما أن القياسات $(x_1 x_2 \dots x_n)$ معلومة فإنه يمكننا كتابة ذلك التوزيع المشترك على الشكل التالي :

$$L(\theta) = f(x_1, \theta) * f(x_2, \theta) \dots f(x_n, \theta) \qquad (92-1)$$

وبافتراض أن المؤشر المجهول θ يأخذ قيماً حقيقية في مجال معلوم، فإن التابع $L(\theta)$ سيأخذ قيماً في مجال معلوم آخر، وسيكون له ضمن ذلك المجال قيمة عظمى (أو صغرى) وتكون مقابلة لقيمة معينة لـ θ . لذلك فإننا سنبحث عن قيمة $L(\theta)$ ونعتبرها تقديراً لـ θ .

ويسمى هذا التقدير بتقدير الامكانية العظمى، لأنه يقابل القيمة العظمى للتوزيع المشترك $L(\theta)$ ، كما يسمى التابع $L(\theta)$ بتابع الامكانية العظمى . وهو عبارة عن التوزيع المشترك لقياسات العينة $(x_1 x_2 \dots x_n)$ ، وقد يحتوي على مؤشر مجهول واحد θ أو أكثر .

ومعلوم أن الشرط اللازم ليأخذ $L(\theta)$ قيمة عظمى هو أن يكون مشتقه الأول معدوماً والثاني سالباً . أي أن يكون:

$$\frac{\partial L(\theta)}{\partial \theta} = 0$$
$$\left.\frac{\partial^2 L(\theta)}{\partial \theta^2}\right|_{\tilde{\theta}} < 0 \qquad (93-1)$$

حيث $\tilde{\theta}$ هي قيمة θ المحسوبة من معادلة المشتق الأول (1−93) .

ولكن بسبب الصيغ الأسية أو الجدائية لمعظم قوانين التوزيع المشتركة، فإننا نستخدم اللوغاريتم الطبيعي $\ln L(\theta)$ بدلاً عن $L(\theta)$ نفسه . وذلك لأن التحويل يسهل علينا كثيراً من الحسابات في التطبيقات العملية , ولا يؤثر على حساب قيمة θ التي نبحث عنها . وبذلك فإن الشرط اللازم ليأخذ $\ln L(\theta)$ قيمة عظمى أيضاً هو أن يكون :

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0 \qquad (94-1)$$
$$\left.\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right|_{\tilde{\theta}} < 0$$

حيث θ̃ هي قيمة θ المحسوبة من المعادلة (1−94) .

ب− الحالة التي يكون لدينا k مؤشراً مجهولاً $(\theta_1\theta_2 \dots \theta_k)$

وعندها فإن قانون التوزيع الاحتمالي لـ X سيتضمن هذه المؤشرات المجهولات، وإن التوزيع الاحتمالي المشترك لقياسات العينة $(x_1x_2 \dots x_n)$ في هذه الحالة يأخذ الشكل التالي :

$$L(X\ \ \theta_1\theta_2 \dots \theta_k) = \prod_{i=1}^{n} f(x_i\,, \theta_1\theta_2 \dots \theta_k) \qquad (\mathbf{95-1})$$

وإن الشروط اللازمة حتى يأخذ التابع المتعدد $L(\theta_1\theta_2 \dots \theta_k)$ قيمة عظمى موضعية هو أن تكون مشتقاته الجزئية الأولى معدومة وإن تكون مصفوفة مشتقاته الجزئية الثانية محددة سلبياً . وهذا يعني أن تتحقق الشروط التالية:

$$\frac{\partial \ln L(\theta_1\theta_2 \dots \theta_k)}{\partial \theta_1} = 0$$
$$\frac{\partial \ln L(\theta_1\theta_2 \dots \theta_k)}{\partial \theta_2} = 0 \qquad (\mathbf{96-1})$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\frac{\partial \ln L(\theta_1\theta_2 \dots \theta_k)}{\partial \theta_k} = 0$$

وبحل هذه المعادلات نحصل منها على التقديرات $(\tilde{\theta}_1\tilde{\theta}_2 \dots \tilde{\theta}_k)$ المطلوبة .

وللتحقق من أن هذه التقديرات تقابل قيمة عظمى موضعية للتابع $\ln L(\theta_1\theta_2 \dots \theta_k)$ يجب أن نتأكد من أن مصفوفة مشتقاته الجزئية الثانية في تلك النقطة $(\tilde{\theta}_1\tilde{\theta}_2 \dots \tilde{\theta}_k)$ محددة سلبياً . وهذا يعني أن تكون قيمة المشتق الثاني بالنسبة للمؤشر الأول سالبة . وأن تكون قيمة المحددات ذات المراتب المتصاعدة متناوبة بالإشارة . ولتوضيح ذلك نرمز لمصفوفة المشتقات الثانية للتابع $(\ln L)$ بالرمز Δ ونكتبها كما يلي:

$$\Delta=\begin{bmatrix} \frac{\partial^2 \ln L}{\partial \theta_1^2} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} & \cdots\cdots & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_2^2} & \cdots\cdots & \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \cdots\cdots & \vdots \\ \frac{\partial^2 \ln L}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_k \partial \theta_2} & \cdots\cdots & \frac{\partial^2 \ln L}{\partial \theta_k^2} \end{bmatrix}_{(\tilde{\theta}_1 \tilde{\theta}_2 \dots \tilde{\theta}_k)} \qquad \mathbf{(97-1)}$$

ومن الواضح أن المصفوفة Δ هي مصفوفة مربعة من المرتبة $(K * K)$ ومتناظرة . وهنا نشير إلى أنه حتى تكون المصفوفة Δ محددة سلبياً في النقطة $\tilde{\theta}(\tilde{\theta}_1 \tilde{\theta}_2 \dots \tilde{\theta}_k)$, يجب أن تحقق عناصرها ومحدداتها الشروط الآتية اللازمة ليأخذ التابع $(\ln L)$ قيمة عظمى وهي :

$$\Delta_1=\left|\frac{\partial^2 \ln L}{\partial \theta_1^2}\right|_{\tilde{\theta}} < 0 \qquad \mathbf{(98-1)}$$

$$\Delta_2=\begin{vmatrix} \frac{\partial^2 \ln L}{\partial \theta_1^2} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} \\ \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_2^2} \end{vmatrix}_{\tilde{\theta}} > 0 \qquad \mathbf{(99-1)}$$

$$\Delta_3=\begin{vmatrix} \frac{\partial^2 \ln L}{\partial \theta_1^2} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_3} \\ \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_2^2} & \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_3} \\ \frac{\partial^2 \ln L}{\partial \theta_3 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_3 \partial \theta_2} & \frac{\partial^2 \ln L}{\partial \theta_3^2} \end{vmatrix}_{\tilde{\theta}} < 0 \qquad \mathbf{(100-1)}$$

وهكذا تتناوب قيم المحددات ذات المراتب المتصاعدة بالإشارة حتى تصل إلى المحدد ذي المرتبة K وهي تشكل k شرطاً مؤلفاً من k محدداً تتزايد بمراتبها وتتناوب بإشاراتها عند النقطة $\tilde{\theta}$.

فإذا كانت الشروط (1−98−99−100) محققة عند نقطة التقديرات $(\tilde{\theta}_1 \tilde{\theta}_2 \dots \tilde{\theta}_k)$ فإن تلك النقطة تقابل قيمة عظمى للتابع $\ln L(\theta)$ أو للتابع $L(\theta)$.

أما إذا كانت جميع قيم تلك الشروط موجبة بما فيها قيمة الشرط الأول− فإن نقطة التقديرات $(\tilde{\theta}_1 \tilde{\theta}_2 \dots \tilde{\theta}_k)$ التي حصلنا عليها من المعادلات (1−96) تقابل قيمة صغرى للتابع $\ln L(\theta)$ أو للتابع $L(\theta)$ وهذا ليس ما نبحث عنه .

أما عندما يكون الشرط الثاني سالباً أو معدوماً فإن نقطة التقديرات $(\tilde{\theta}_1 \tilde{\theta}_2 \dots \tilde{\theta}_k)$ تقابل قيمة حرجة للتابع $\ln L(\theta)$ ، فهي ليست عظمى وليست صغرى بل تمثل نقطة وسطى كالنقطة التي تمثل مركز سرج الحصان ولذاك تسمى نقطة سرجية .

وأخيراً نشير إلى مصفوفة المشتقات الجزئية الثانية مرتبطة بمصفوفة التباينات المختلفة للتقديرات $(\tilde{\theta}_1 \tilde{\theta}_2 \dots \tilde{\theta}_k)$. وإن العلاقة بينهما هي:

$$V = E[-\Delta^{-1}] = \begin{bmatrix} \sigma^2_{\theta_1} & Cov(\theta_1\,\theta_2) & \dots\dots & Cov(\theta_1\,\theta_k) \\ Cov(\theta_2\,\theta_1) & \sigma^2_{\theta_2} & \dots\dots & Cov(\theta_2\,\theta_k) \\ \vdots & \vdots & \dots\dots & \vdots \\ Cov(\theta_k\,\theta_1) & Cov(\theta_k\,\theta_2) & \dots\dots & \sigma^2_{\theta_k} \end{bmatrix} \quad (101-1)$$

ويستفاد من هذه العلاقة في حساب المصفوفة V من خلال مصفوفة المشتقات الثانية للوغاريتم التوزيع المشترك.

مثال1−7: عن التوزيعات التي تتضمن وسيطين مجهولين μ و σ^2 :

لنفترض أن X يخضع للتوزيع الطبيعي $N(\mu\,,\sigma^2)$ حيث أن μ وσ^2 مجهولين ويطلب تقديرهما وحساب الخطأ المعياري المرتكب في تقدير كل منها, ولذلك نأخذ عينة من قيم X المستقلة هي: $(x_1 x_2\; x_3 \dots x_n)$ ونحسب متوسطها وتباينها من العلاقتين :

$$\bar{x} = \frac{1}{n}\sum x_i \quad (102-1)$$

$$D^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 \quad (103-1)$$

ثم نقوم بتشكيل التوزيع المشترك لقيم X المستقلة فنجد أن :

$$L(X_1 X_2\; X_3 \dots X_n\,,\mu\,,\sigma^2) = \frac{1}{\left(\sqrt{2\pi} * \sigma\right)^n} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2} \quad (104-1)$$

ثم نأخذ اللوغاريتم الطبيعي للطرفين فنحصل بعد تحويل σ إلى $\sigma = \sqrt{\sigma^2}$ لضرورات الاشتقاق , على أن:

$$\ln L(\mu\,,\sigma^2) = -n\ln\left(\sqrt{2\pi} * \sqrt{\sigma^2}\right) - \frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2 \quad \mathbf{(105-1)}$$

$$\ln L(\mu\,,\sigma^2) = -n\ln\left(\sqrt{2\pi}\right) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \quad \mathbf{(106-1)}$$

ثم نأخذ المشتقين الجزئيين بالنسبة لـ μ و(σ^2) ونضعهما مساويين للصفر فنحصل على المعادلتين التاليتين :

$$\frac{\partial \ln L(\mu\,,\sigma^2)}{\partial\mu} = -\frac{1}{2\sigma^2}\sum[-2(x_i-\mu)] = 0 \quad \mathbf{(107-1)}$$

$$\frac{\partial \ln L(\mu\,,\sigma^2)}{\partial(\sigma^2)} = -\frac{n}{2} * \frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum(x_i-\mu)^2 = 0 \quad \mathbf{(108-1)}$$

ومن المعادلة (1−107) الأولى نجد أنه يمكننا حساب تقدير μ من العلاقة التالية:

$$\tilde{\mu} = \frac{1}{n}\sum x_i = \bar{x} \quad \mathbf{(109-1)}$$

ومن المعادلة (1−108) وبعد تعويض μ بـ $\bar{x}$ **يمكننا الحصول على تقدير لـ σ^2 من العلاقة التالية :**

$$\tilde{\sigma}^2 = \frac{1}{n}\sum(x_i-\bar{x})^2 = D^2 \quad \mathbf{(110-1)}$$

ولكنه تقدير متحيز للتباين σ^2 كما رأينا سابقاً.

وحتى يكون هذان التقديران $(\bar{x}, D^2)$ مقابلين لقيمة عظمى للتوزيع المشترك $L(\mu, \sigma^2)$, فإنه يجب أن يكون فيهما قيمة **المشتق الثاني للمؤشر الأول عند تلك النقطة سالبة** , وأن تكون قيمة محدد مصفوفة المشتقات الثانية موجبة , أي يجب أن يكون: $\left.\frac{\partial^2 \ln L(\mu,\sigma^2)}{\partial\mu^2}\right|_{\substack{\tilde{\mu}=\bar{x}\\ \sigma^2=D^2}} < 0$

$$\left|\begin{matrix}\frac{\partial^2 \ln L(\mu,\sigma^2)}{\partial\mu^2} & \frac{\partial^2 \ln L(\mu,\sigma^2)}{\partial\mu\partial(\sigma^2)} \\ \frac{\partial^2 \ln L(\mu,\sigma^2)}{\partial(\sigma^2)*\partial\mu} & \frac{\partial^2 \ln L(\mu,\sigma^2)}{\partial(\sigma^2)^2}\end{matrix}\right|_{\substack{\tilde{\mu}=\bar{x}\\ \sigma^2=D^2}} > 0 \qquad (111-1)$$

لذلك نقوم بحساب هذه المشتقات فنجد أنها تساوي:

$$\begin{aligned}\frac{\partial^2 \ln L(\mu,\sigma^2)}{\partial\mu^2} &= \frac{\partial}{\partial\mu}\left[+\frac{1}{\sigma^2}\sum(x_i-\mu)\right]_{\substack{\tilde{\mu}=\bar{x}\\ \sigma^2=D^2}} \\ &= \frac{1}{\sigma^2}\sum(-1)\Big|_{\substack{\tilde{\mu}=\bar{x}\\ \sigma^2=D^2}} = \frac{-n}{D^2} < 0\end{aligned} \qquad (\mathbf{112-1})$$

وهو يأخذ قيمة سالبة عند التقديرين $(\bar{x}, D^2)$ وتساوي $\left(\frac{-n}{D^2}\right)$, وكذلك نجد أن المشتقات الأخرى تساوي:

$$\begin{aligned}\frac{\partial^2 \ln L(\mu,\sigma^2)}{\partial(\sigma^2)^2} &= \frac{n}{2}\frac{1}{\sigma^4} - \frac{1}{\sigma^6} + \sum_{i=1}^{n}(x_i-\mu)^2\Big|_{\substack{\tilde{\mu}=\bar{x}\\ \sigma^2=D^2}} = \\ &= \frac{n}{2D^4} - \frac{nD^2}{D^6} = \frac{-n}{2D^4}\end{aligned} \qquad (\mathbf{113-1})$$

$$\frac{\partial^2 \ln L(\mu,\sigma^2)}{\partial\mu\partial(\sigma^2)^2} = \frac{\partial^2 \ln L(\mu,\sigma^2)}{\partial(\sigma^2)^2\partial\mu} = 0 \qquad (\mathbf{114-1})$$

وبذلك نجد أن مصفوفة المشتقات الثانية تأخذ عند هذين التقديرين الشكل التالي:

$$\Delta_2 = \begin{bmatrix}\frac{-n}{D^2} & 0 \\ 0 & \frac{-n}{2D^4}\end{bmatrix} \qquad (115-1)$$

وإن قيمة محدد هذه المصفوفة تساوي:

$$|\Delta_2| = \frac{+n^2}{2D^6} > 0 \qquad (116-1)$$

وهذا يعني أن التقديرين $(\bar{x}, D^2)$ يقابلان قيمة عظمى للتوزيع المشترك $L(\mu, \sigma^2)$.

ثم نقوم بحساب مقلوب هذه المصفوفة فنجد أن:

$$\Delta_2^{-1} = \frac{2D^6}{n^2}\begin{bmatrix} \frac{-n}{2D^4} & 0 \\ 0 & \frac{-n}{D^2} \end{bmatrix} = \begin{bmatrix} \frac{-D^2}{n} & 0 \\ 0 & \frac{-2D^4}{n} \end{bmatrix} \qquad (117-1)$$

ولإيجاد تقدير تباين كل من $\bar{x}$ و D^2 نقوم بحساب التوقع الرياضي للمصفوفة $[-\Delta_2^{-1}]$ فنحصل على مصفوفة التباينات المقابلة لكل منهما كما يلي :

$$V = E[-\Delta_2^{-1}] = E\begin{bmatrix} \frac{D^2}{n} & 0 \\ 0 & \frac{2D^4}{n} \end{bmatrix} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix} \qquad (118-1)$$

ومنها نستنتج أن تباين التقدير $\bar{x}$ لـ μ يساوي: $Var(\bar{x}) = \frac{\sigma^2}{n}$

وأن تباين التقدير D^2 لـ σ^2 يساوي: $Var(D^2) = \frac{2\sigma^4}{n}$

وكذلك نجد أن التباين المشترك لـ $\bar{x}$ و D^2 يساوي الصفر أي أن:

$$Cov(\bar{x}, D^2) = Cov(D^2, \bar{x}) = 0 \qquad (119-1)$$

وهذا يعني أن التقدير $\bar{x}$ و D^2 غير مرتبطين وبما أنها يخضعان للتوزيع الطبيعي فإنهما مستقلان .

تمارين الفصل الأول

1-إذا كان $\tilde{\theta}$ تقديراً غير متحيز ومتسقاً لـ θ فبرهن على

- أن $\tilde{\theta}^2$ هو تقدير متحيز لـ θ^2 ؟
- أن $\tilde{\theta}^2$ هو تقدير متسق لـ θ^2 ؟

2-أوجد تقدير الإمكانية العظمى $\tilde{\theta}$ للمؤشر θ في كل من التوزيعات التالية :

$$f(x,\theta)=\theta^x(1-\theta)^{1-x} \quad (x \text{ صحيح غير سالب})$$

$$f(x,\theta)=\frac{1}{2}e^{-|x-\theta|} \quad (-\infty<x<\infty)$$

$$f(x,\theta)=e^{-(x-\theta)} \quad (\theta<x<\infty)$$

$$f(x,\theta)=\frac{x^{p-1}e^{-x/\theta}}{\Gamma(p)*\theta^p} \quad x\geq 0$$

$$f(x,\theta)=\frac{\theta^x e^{-\theta}}{x!} \quad (x \text{ صحيح موجب})$$

ثم ادرس عدم تحيز كل منها ثم فعاليته واتساقه ؟

3- أوجد بطريقة العزوم تقدير θ في التوزيع ؟

$$f(x,\theta)=(1+\theta)x^\theta \quad (0\leq x\leq 1)$$

ثم أوجد تقدير الإمكانية العظمى لـ θ ؟

4- أوجد بطريقة العزوم تقديراً لكل من a و b في التوزيع :

$$f(x,a,b)=\frac{1}{b-a} \quad (a<x<b)$$

ثم ادرس عدم تحيز تقديريهما واستخلص من ذلك تقديراً غير متحيز لكل منهما .

5- لنفترض أننا سحبنا عينتين بحجمين n_1 و n_2 من مجتمعين طبيعيين خاضعين لـ $N(\mu_1,\sigma_1^2)$ و $N(\mu_2,\sigma_2^2)$ على الترتيب .

أوجد تقدير الامكانية العظمى للفرق $\mu_1-\mu_2$.

وبفرض أن حجم العينة الكلية $n=n_1+n_2$ ثابت، فكيف يمكن أن نجزئ n بحيث يكون تباين $\mu_1-\mu_2$ أصغر ما يمكن .

6-ليكن لدينا التوزيع الطبيعي لـ (x,y) التالي:

$$f(x,y)=\frac{1}{\sigma_x\sigma_y 2\pi\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}A}$$

حيث أن:

$$A = \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2r\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]$$

وحيث أن: $1 > |r|$ و $0 < \sigma_x , \sigma_y$ و$\infty > x , y > -\infty$

والمطلوب إيجاد تقدير الإمكانية العظمى لكل من المؤشرات التالية:

1- للمؤشرات r و σ_x^2 و σ_y^2 عندما يكون μ_x و μ_y معلومين .

2- للمؤشرات الخمسة r و σ_x^2 و σ_y^2 و μ_x و μ_y .

3- اوجد مصفوفة التباينات التقاربية لكل من الحالتين السابقتين .

الفصل الثاني : التقدير المجالي

2-1: تمهيد:

عندما كنا نقوم بتقدير المؤشرات الإحصائية في المجتمع بوساطة ما يقابلها في العينة المسحوبة من ذلك المجتمع فإننا كنا نحصل على تقديرات نقطية لتلك المؤشرات . وبالرغم من أن معظم تلك المؤشرات هي تقديرات غير متحيزة ومتماسكة وفعالة، ألا أن أياً منها لا يعطينا أي درجة من الثقة فيه .

وهكذا يتضح لنا أنه لابد لنا من إيجاد وسيلة أخرى تؤكد لنا, أو على الأقل تضمن لنا , أن تكون تلك التقديرات غير بعيدة عن الحقيقة والواقع وباحتمال معين وموثوق .

إن الوسيلة المستخدمة في البحث عن تقديرات موثوقة باحتمال معين تسمى بمجالات الثقة للمؤشر المقدر . ومجال الثقة هو مجال يشترط فيه أن يضمن لنا باحتمال كبير ومحدد، أن تقع القيمة الحقيقية للمؤشر الذي حصلنا على تقديره, في ذلك المجال .

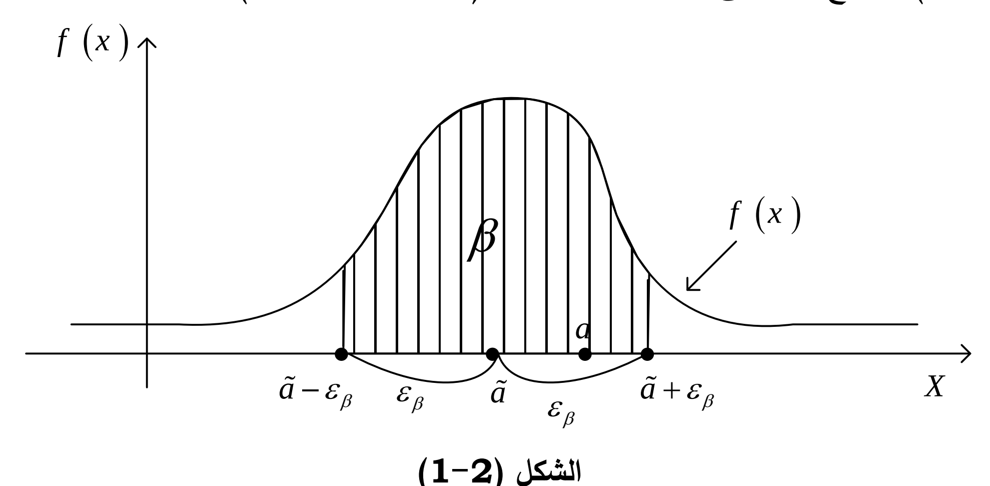
فإذا كان θ مؤشراً في المجتمع وكان $\tilde{\theta}$ تقديره من العينة فكيف يمكن أن ننشئ مجالاً يحوي المقدار θ باحتمال معلوم قدره β ؟

لإنشاء أي مجال يجب أن نبحث عن عنصرين هما: إما مركزه ونصف طوله أو قيمتي حدية C_1 و C_2 $(C_1 < C_2)$ وبما أن المقدار θ هو قيمة عددية مجهولة في المجتمع، فإننا إذا جعلنا تقديره $\tilde{\theta}$ مركزاً للمجال المطلوب وافترضنا أن نصف طوله ε_β فإن ε_β يجب أن يحقق الشرط التالي:

$$P\left(\tilde{\theta} - \varepsilon_\beta < \theta < \tilde{\theta} + \varepsilon_\beta\right) = \beta \qquad (1-2)$$

حيث أن β هو احتمال الثقة وحدد مسبقاً .

وإن الشكل (2-1) يوضح لنا معنى مجال الثقة المذكور(تستبدل a بـ θ).



الشكل (2-1)

ومن الواضح أن ε_β تابع للاحتمال β ويجب حسابه بحيث يكون الشرط (2-1) محققاً .

إن طريقة حساب ε_β تختلف من حالة لأخرى . ولكن بصورة عامة، حتى نستطيع حساب ε_β يجب أن نقوم بتحويل لأطراف المتراجحتين في (2-1) بحيث نحصل في الوسط على متحول عشوائي ذي توزيع معروف .

وأن أشهر هذه التحويلات هو تحويلها إلى الشكل الذي يعطينا متحولاً معيارياً ذا توزيع معروف من الشكل التالي: $x=\frac{\tilde{\theta}-\theta}{\sigma_{\tilde{\theta}}}$:

حيث $\sigma_{\tilde{\theta}}$ هو الانحراف المعياري للتقدير $\tilde{\theta}$.

وإذا طرحنا $\tilde{\theta}$ من أطراف المتراجحة (2−1) ثم ضربناها بـ (−1), وقسمناها على $\sigma_{\tilde{\theta}}$, فإننا نحصل على الشرط المكافئ التالي:

$$P\left(-\frac{\varepsilon_\beta}{\sigma_{\tilde{\theta}}}<\frac{\tilde{\theta}-\theta}{\sigma_{\tilde{\theta}}}<\frac{\varepsilon_\beta}{\sigma_{\tilde{\theta}}}\right)=P\left(-\frac{\varepsilon_\beta}{\sigma_{\tilde{\theta}}}<\mathrm{x}<\frac{\varepsilon_\beta}{\sigma_{\tilde{\theta}}}\right)=\beta \qquad (2-2)$$

وبفرض أن المتحول $x=\frac{\tilde{\theta}-\theta}{\sigma_{\tilde{\theta}}}$ هو متحول عشوائي مستمر ويخضع لقانون التوزيع الاحتمالي $f(x)$, نجد أن الشرط المطلوب(2−2) يصبح مساوياً لما يلي:

$$\int_{-\frac{\varepsilon_\beta}{\sigma_{\tilde{\theta}}}}^{\frac{\varepsilon_\beta}{\sigma_{\tilde{\theta}}}} f(x)dx=\beta \qquad (3-2)$$

وبحل هذه المعادلة نحصل على نصف الطول ε_β , كما سنرى في الفقرات اللاحقة.

وسنعالج مسألة إنشاء مجال ثقة لعدد من المؤشرات هي: المتوسط والاجمالي والنسبة ...الخ .

2−2: إنشاء مجال ثقة لمتوسط مجتمع طبيعي μ (أو توقعه μ)

نعلم أن متوسط المجتمع μ يقدر بوساطة متوسط العينة $\bar{x}$ وأن إنشاء مجال ثقة باحتمال قدره β لـ μ حول $\bar{x}$ يقتضي حساب ε_β الذي يحقق الشرط (2−1) التالي:

$$P(\bar{x}-\varepsilon_\beta<\mu<\bar{x}+\varepsilon_\beta)=\beta \qquad (4-2)$$

وبطرح $\bar{x}$ من أطراف المتراجحة ثم ضربها بـ (−1) وتقسيمها على $\sigma_{\bar{x}}$ يتحول الشرط (2−4) إلى الشكل التالي :

$$P\left(-\frac{\varepsilon_\beta}{\sigma_{\bar{x}}}<\frac{\bar{x}-\mu}{\sigma_{\bar{x}}}<\frac{\varepsilon_\beta}{\sigma_{\bar{x}}}\right)=\beta=1-\alpha \qquad (5-2)$$

حيث أن α هو مستوى الدلالة وهو يعبر عن الاحتمال الذي يجب تركه خارج مجال الثقة المطلوب .

إن حساب ε_β يقتضي أن يكون $\sigma_{\bar{x}}$ معلوماً أو مقدراً . وبما أن توزيع المتحول $\frac{\bar{x}-\mu}{\sigma_{\bar{x}}}$, يختلف حسب ما يكون $\sigma_{\bar{x}}$ معلوماً أو مقدراً , لذلك فإننا سنميز بين الحالتين التاليتين:

2−2−1: الحالة التي يكون فيها $\sigma_{\bar{x}}$ مقدراً : (أو عندما يكون حجم العينة n صغيراً) :

نعلم أن $\sigma_{\bar{x}}^2$ في حالة السحب مع الإعادة يساوي $\sigma_{\bar{x}}^2=\frac{\sigma^2}{n}$, وهو يقدر بوساطة العلاقة $\tilde{\sigma}_{\bar{x}}^2=\frac{S^2}{n}$

حيث أن s^2 هو تباين العينة المصحح ويحسب من العلاقة التالية:

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

وبذلك نجد أن الشرط (2-5) يأخذ الشكل التالي:

$$P\left(\frac{-\varepsilon_\beta}{s/\sqrt{n}} < \frac{\bar{x}-\mu}{s/\sqrt{n}} < \frac{\varepsilon_\beta}{s/\sqrt{n}}\right) = \beta = 1-\alpha \qquad (6-2)$$

وبملاحظة أن المتحول الأوسط $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ يخضع لتوزيع ستودينت $T_{(n-1)}$ ذي $(n-1)$ درجة حرية .

لذلك نقوم بحساب الاحتمال الأيسر من التوزيع $T_{(n-1)}$ المتناظر بالنسبة إلى المحور الشاقولي فنجد أن :

$$\int_{\frac{-\varepsilon_\beta}{s/\sqrt{n}}}^{\frac{\varepsilon_\beta}{s/\sqrt{n}}} T(t)_{(n-1)}dt = 2\int_{0}^{\frac{\varepsilon_\beta}{s/\sqrt{n}}} T(t)_{(n-1)}dt = 1-\alpha$$

وهذا يكافئ :

$$\left[P\left(t \le \frac{\varepsilon_\beta}{s/\sqrt{n}}\right) - P(t \le 0)\right] = \frac{1-\alpha}{2}$$

$$P\left(t \le \frac{\varepsilon_\beta}{s/\sqrt{n}}\right) - \frac{1}{2} = \frac{1}{2} - \frac{\alpha}{2}$$

$$P\left(t \le \frac{\varepsilon_\beta}{s/\sqrt{n}}\right) = 1 - \frac{\alpha}{2} \qquad (7-2)$$

وبمقارنة هذه العلاقة مع العلاقة التي تعطينا قيم t_p الحدية في الجداول، التي لها الشكل اليساري التالي:

$$P(t < t_p) = \int_{-\infty}^{t_p} T(t)(_{(n-1)}dt = P \qquad (8-2)$$

وبمقارنة العلاقتين (2-7)و(2-8), فإننا نجد أن*:

$$\frac{\varepsilon_\beta}{s/\sqrt{n}} = t_{\left(1-\frac{\alpha}{2},n-1\right)} \qquad (10-2)$$

حيث أن $t_{\left(1-\frac{\alpha}{2},n-1\right)}$ هي قيمة متحول (ستودينت) التي تقابل الاحتمال اليساري $\left(1-\frac{\alpha}{2}\right)$ ودرج الحرية $(n-1)$, وهي تؤخذ من الجدول الملحق بآخر الكتاب, والشكل (2-2) يوضح لنا معنى الاحتمال الوارد في (2-8) بالنسبة إلى توزيع (ستودينت) .

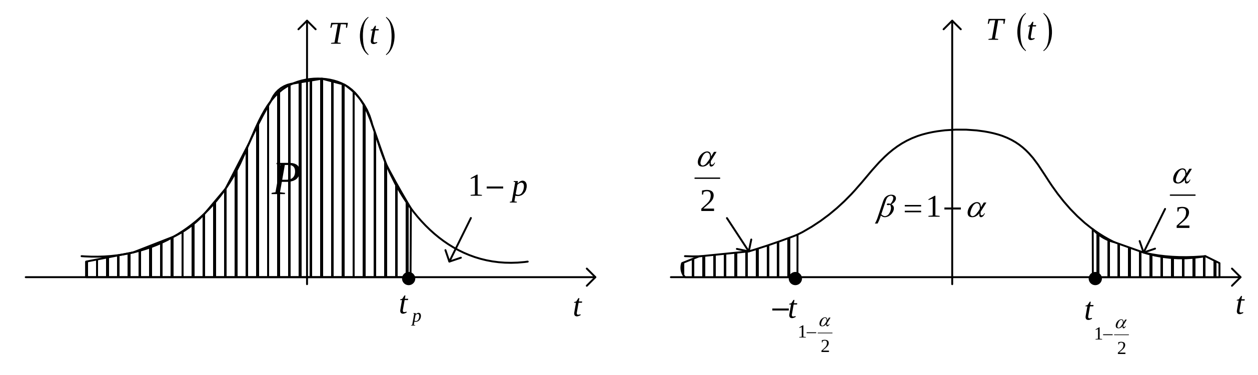
---

* إذا كان جدول (ستودينت) يعتمد على العلاقة اليمينية التالية:

$$P(t > t'_q) = \int_{t'_q}^{+\infty} T(t)_{(n-1)}dt = q \qquad (9-2)$$

فإننا سنجد أن : $\frac{\varepsilon_\beta}{s/\sqrt{n}} = t'_{\frac{\alpha}{2},n-1}$. وهنا نشير إلى أن $t_{1-\frac{\alpha}{2}} = t'_{\frac{\alpha}{2}}$

وذلك لأنهما يمثلان النقطة الفاصلة بين الاحتمالين P وq (حيث p+q=1) . وكذلك الأمر بالنسبة إلى التوزيعين χ^2 و F



الشكل (2-2)

وبذلك نكون قد وجدنا أن نصف طول المجال المطلوب ε_β يساوي :

$$\varepsilon_\beta = t_{\left(1-\frac{\alpha}{2},n-1\right)} * s/\sqrt{n} \qquad (11-2)$$

وبالتعويض في العلاقة (2-6) نحصل على أن:

$$P\left(-t_{\left(1-\frac{\alpha}{2}\right)} < \frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{\left(1-\frac{\alpha}{2}\right)}\right) = 1-\alpha \qquad (12-2)$$

أو في العلاقة (2-4) فنجد أن مجال الثقة المطلوب لـ μ هو:

$$P\left(\bar{x} - t_{1-\frac{\alpha}{2}} * s/\sqrt{n} < \mu < \bar{x} + t_{-\frac{\alpha}{2}} * s/\sqrt{n}\right) = 1-\alpha \qquad (13-2)$$

وأخيراً نشير إلى أن نصف طول المجال ε_β يعد متحولاً عشوائياً وتابعاً لمعطيات العينة $\left(s \text{ و } \sqrt{n}\right)$ كما أن ε_β يتناقص كلما ازداد حجم العينة n, أي أن مجال الثقة يضيق كلما زاد حجم العينة n .

فإذا كان حجم العينة n صغيراً أو كان التباين s^2 كبيراً (حالة مجتمع غير متجانس أو غير طبيعي) فإننا قد نحصل على مجال ثقة عريض, وقد يكون مجالاً غير ذي جدوى وذلك بصرف النظر عن احتمال الثقة المقابل له . كأن نحصل على مجال ثقة لطول الإنسان مساوياً للمجال $[0\ ,\ 300\ cm]$ وباحتمال قدره $\beta = 1$, لأن مجالاً كهذا غير ذي جدوى بغض النظر عن احتمال الثقة الكامل المقابل له. ولهذا فإن حجم العينة يجب أن يكون كبيراً إلى الحد الذي يؤمن لنا، وباحتمال كبير، عرضاً لمجال الثقة صغيراً إلى الحد الذي يجعله مقبولاً في التطبيقات العملية . وهذا ما نعالجه في نظرية العينات .

- ملاحظة: إذا كان السحب قد جرى بدون إعادة فإن تقدير $\tilde{\sigma}_{\bar{x}}$ يساوي :

$$\tilde{\sigma}_{\bar{x}} = \sqrt{\frac{N-n}{N} * \frac{s^2}{n}}$$

ولكن بما أن:

$$\sqrt{\frac{N-n}{N} * \frac{s^2}{n}} < \sqrt{\frac{s^2}{n}}$$

فإننا سنحصل على مجال ثقة أصغر من مجال الثقة للسحب مع الإعادة, ولكن إذا كان حجم العينة كبيراً فيفضل الابقاء على التقدير $s/\sqrt{n}$ في كلتا حالتي السحب, حيث أننا نحصل على مجال ثقة أعرض من

مجال الثقة المطلوب ويقابل احتمالاً أكبر من β المحددة , أما إذا كان ذلك سيؤدي إلى الحصول على مجال ثقة عريض جداً فإننا نستخدم التقدير الدقيق لحالة السحب بدون إعادة التالي:

$$\tilde{\sigma}_{\bar{x}} = \sqrt{\frac{N-n}{N} * \frac{s^2}{n}}$$

2-2-2: الحالة التي يكون فيها $\sigma_{\bar{x}}$ معلوماً (أو عندما يكون حجم العينة n كبيراً) .

في الحقيقة أن $\sigma_{\bar{x}}$ لا يكون معلوماً إلا إذا كانت العينة شاملة لجميع عناصر المجتمع . ولكن عندما يكون حجم العينة كبيراً فإنه يمكننا اعتبار تباين العينة s^2 قيمة تقريبية لتباين المجتمع σ^2 . وبما أن $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ (في حالة السحب مع الإعادة أو عندما يكون n كبيراً), فإن المقدار $\sigma_{\bar{x}}^2$ يكون قد أصبح قيمة عددية وليس متحولاً عشوائياً وتأخذ العلاقة (2-5) الشكل التالي:

$$P\left(\frac{-\varepsilon_\beta}{\sigma/\sqrt{n}} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{\varepsilon_\beta}{\sigma/\sqrt{n}}\right) = \beta = 1-\alpha \qquad (14-2)$$

وبما أن المتحول $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ يخضع للتوزيع الطبيعي المعياري $N(0,1)$ فإن الاحتمال في الطرف الأيسر يساوي:

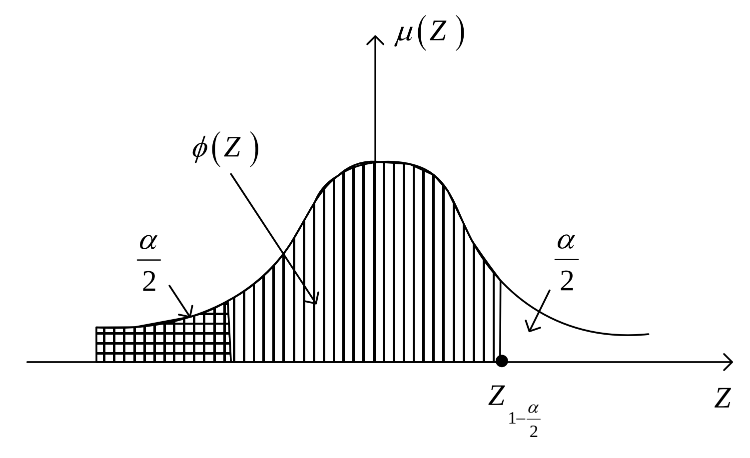
$$P\left(\frac{-\varepsilon_\beta}{\sigma/\sqrt{n}} < z < \frac{\varepsilon_\beta}{\sigma/\sqrt{n}}\right) = \phi\left(\frac{\varepsilon_\beta}{\sigma/\sqrt{n}}\right) - \phi\left(\frac{-\varepsilon_\beta}{\sigma/\sqrt{n}}\right) = 1-\alpha \qquad (15-2)$$

ومنها نجد أن:

$$2\phi\left(\frac{\varepsilon_\beta}{\sigma/\sqrt{n}}\right) - 1 = 1-\alpha \qquad (16-2)$$

$$\frac{\varepsilon_\beta}{\sigma/\sqrt{n}} = \phi^{-1}\left(1-\frac{\alpha}{2}\right) = z_{1-\frac{\alpha}{2}} \qquad (17-2)$$

حيث رمزنا بـ $z_{1-\frac{\alpha}{2}}$ إلى قيمة المتحول المعياري z التي تقابل الاحتمال الأيسر$\left(1-\frac{\alpha}{2}\right)$, وهي تؤخذ من جدول التوزيع الطبيعي الملحق بآخر الكتاب , والشكل رقم (2-3) يوضح لنا معنى الاحتمال الوارد في (2-15) بالنسبة إلى التوزيع الطبيعي المعياري .



الشكل (2-3)

وبذلك نحصل على نصف طول المجال ε_β ويساوي:

$$\varepsilon_\beta = z_{1-\frac{\alpha}{2}} * \sigma/\sqrt{n} \qquad (18-2)$$

وباستبدال σ بقيمته التقريبية s نحصل على نصف الطول ε_β من العلاقة:

$$\varepsilon_\beta = z_{1-\frac{\alpha}{2}} * s/\sqrt{n} \qquad (19-2)$$

وبالتعويض في العلاقة (2−14) نجد أن:

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{x}-\mu}{s/\sqrt{n}} < z_{1-\frac{\alpha}{2}}\right) = 1-a \qquad (20-2)$$

وبالتعويض في (2−4) نحصل على مجال الثقة لتوقع المجتمع μ التالي:

$$P\left(\bar{x} - z_{1-\frac{\alpha}{2}} * s/\sqrt{n} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}} * s/\sqrt{n}\right) = 1-\alpha \qquad (21-2)$$

وهو المجال المطلوب .

- **مثال (2−1)**: لنفترض أننا سحبناعينة بحجم $n = 10$ من الطلاب فكانت قياسات أطوالهم كما يلي:

$$X: 170 , 161 , 169 , 173 , 160 , 174 , 165 , 162 , 167 , 164\ cm$$

والمطلوب إيجاد تقدير لمتوسط طول الطالب تم إيجاد مجال ثقة له باحتمال ثقة قدره $\beta = 0,95$

الحل: بما ان بحثنا لا يشمل كل الطلاب إذن فإن المتوسط الحقيقي لطول الطالب أوتوقعه μ سيبقى مجهولاً .

لذلك نقوم بتقديره عن طريق معطيات العينة: فنجد أن:

$$\bar{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1665}{10} = 166,5\ cm$$

ولإيجاد مجال الثقة نحتاج لحساب تباين العينة s^2 وهو يساوي:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{218,5}{9} = 24,28$$

$$s = \sqrt{24,28} = 4,927$$

وبما أن $\beta = 0,95$ فإن مستوى الدلالة يساوي : $\alpha = 1 - \beta = 0,95$, وبما أن حجم العينة n صغير جداً$(n < 30)$, فإننا نكون أمام الحالة الأولى، حيث يكون $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ خاضعاً لتوزيع (ستودينت) ويكون نصف طول المجال ε_β مساوياً :

$$\varepsilon_\beta = t_{1-\frac{\alpha}{2}} * s/\sqrt{n}$$

ومن جدول ستودينت نجد أن قيمة المتحول t التي تقابل الاحتمال اليساري $1 - \frac{\alpha}{2} = 0,975$ ودرجة الحرية تساوي: $n - 1 = 9$

$$t_{1-\frac{\alpha}{2},9} = 2,26$$

وبذلك نجد أن مجال الثقة المطلوب يساوي :

$$P\left(\bar{x}-t_{1-\frac{\alpha}{2}}*s/\sqrt{n}<\mu<\bar{x}+t_{1-\frac{\alpha}{2}}*s/\sqrt{n}\right)=1-\alpha$$
$$P\left(166,5-2,26*\frac{4,927}{\sqrt{10}}<\mu<166,5+2,26*\frac{4,927}{\sqrt{10}}\right)=0,95$$
$$P(162,98<\mu<170,02)=0,95$$

أي أنه يمكننا القول أن متوسط طول الطالب μ أو توقعه الرياضي $E(x)$ يقع في المجال [162,98,170,02] باحتمال قدره β = 0,95 وأن اطوال 5%من الطلاب يمكن أن تقع خارج هذا المجال.

- **مثال (2-2)**: لنفترض أننا قمنا بدراسة على 50 طفلاً فور ولادتهم, فوجدنا أن متوسط أطوالهم في هذه العينة كان $\bar{\ell}=53c\ m$ ومتوسط أوزانهم كان $\bar{w}=3,5\ \text{kg}$ وتباين الطول $s_\ell^2=64$ وتباين الوزن كان $s_w^2=25$. فأوجد مجال الثقة لكل من متوسط الطول $\bar{L}$ ومتوسط الوزن $\bar{w}$ وذلك باحتمال قدره β = 0,95 .

الحل: بما ان حجم العينة يعد كبيراً $(n>30)$ إذن فإننا نكون أمام الحالة الثانية، حيث يكون نصف طول مجال الثقة المطلوب مساوياً:

$$\varepsilon_\beta=z_{1-\frac{\alpha}{2}}*s/\sqrt{n}$$

ومن جدول التوزيع الطبيعي المعياري نجد أن:

$$z_{1-\frac{\alpha}{2}}=z_{0,975}=1,96$$

وبذلك نجد أن مجال الثقة المطلوب لمتوسط الطول $\bar{L}$ هو المجال:

$$P\left(\bar{\ell}-z_{1-\frac{\alpha}{2}}\frac{s_\ell}{\sqrt{n}}<\bar{L}<\bar{\ell}+z_{1-\frac{\alpha}{2}}\frac{s_\ell}{\sqrt{n}}\right)=1-\alpha=0,95$$

وبالتعويض نجد أن:

$$P\left(53-1,96\frac{8}{\sqrt{50}}<\bar{L}<53+1,96\frac{8}{\sqrt{50}}\right)=1-\alpha=0,95$$

وبذلك يكون مجال الثقة المطلوب لـ $\bar{L}$ هو المجال [50,78 , 55,2] .

وبطريقة مشابهة نجد مجال الثقة لمتوسط الوزن $\bar{w}$ هو المجال:

$$P\left(\bar{w}-z_{1-\frac{\alpha}{2}}\frac{s_w}{\sqrt{n}}<\bar{w}<\bar{w}+z_{1-\frac{\alpha}{2}}\frac{s_w}{\sqrt{n}}\right)=0,95$$
$$P\left(3,5-1,96\frac{5}{\sqrt{50}}<\bar{w}<3,5+1,96\frac{5}{\sqrt{50}}\right)=0,95$$

وبذلك نجد أن مجال الثقة المطلوب لـ $\bar{w}$ هو المجال [2,114 , 4,886] .

2-3: إنشاء مجال الثقة لنسبة خاصة ما في مجتمع طبيعي:

لنرمز إلى الخاصة المدروسة بـ A ونسبة وجودها في المجتمع بـ R، ولتقدير R نفترض أننا سحبنا عينة عشوائية بحجم n ووجدنا أن نسبة تلك الخاصة في العينة تساوي r . والمطلوب هو إيجاد مجال ثقة حول r بحيث يحتوي على النسبة المجهولة R باحتمال ثقة قدره β ؟ .

إن مسألة إنشاء مجال الثقة للنسبة R تشبه تماماً إنشاء مجال الثقة للمتوسط μ , التي عالجناها في الفقرة السابقة. ولتوضيح ذلك نفترض متحولاً X مرافقاً لكل عنصرمن عناصر العينة ويأخذ أحدى القيمتين التاليتين :

$x_i = 1$ إذا كان العنصر i يتميز بالخاصة المدروسة .

$x_i = 0$ إذا كان العنصر i لا يتميز بالخاصة المدروسة .

فعندئذ نجد أن عدد عناصر العينة التي تتصف بتلك الخاصة يساوي:

$$m = \sum_{i=1}^{n} x_i \qquad (x_i : 0 , 1) \qquad (22-2)$$

وهكذا نجد أن النسبة في العينة تساوي:

$$r = \frac{m}{n} = \frac{\sum x_i}{n} = \bar{x} \qquad (23-2)$$

وهكذا يمكننا اعتبار r وكأنه المتوسط $\bar{x}$ لقيم X التي تأخذ احدى القيمتين 0 أو 1, وكذلك نجد أن تباين العينة s^2 يساوي :

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right)$$
$$= \frac{1}{n-1}\left[\sum x_i^2 - 2\bar{x}\sum x_i + n\,\bar{x}^2\right]$$

وبما أن : $\sum x_i^2 = \sum x_i = m = nr$, إذن نجد أن:

$$s^2 = \frac{1}{n-1}[nr - 2r * nr + nr^2] = \frac{n}{n-1}r(1-r) \qquad (24-2)$$

$$s^2 = \frac{n}{n-1}r\,q \qquad : q = 1 - r \qquad (24-2)$$

ومن ثم نجد أن تباين التقدير r يقدر بالعلاقة التالية :

$$\tilde{\sigma}_r^2 = \frac{s^2}{n} = \frac{r * q}{n-1} \approx \frac{rq}{n} \qquad (25-2)$$

وذلك في حالة السحب مع الإعادة .

أما إذا كان السحب بدون إعادة فإن تقدير σ_r^2 يعطى بالعلاقة :

$$\tilde{\sigma}_r^2 = \frac{N-n}{N}\ \frac{s^2}{n} = \frac{N-n}{N}\ \frac{rq}{n} \qquad (26-2)$$

وعندما يكون حجم العينة كبيراً , فإننا سنتبنى التقدير الوارد في (2−25) في كلتا حالتي السحب، إلا إذا أشرنا إلى عكس ذلك . وبصورة عامة فإن مسالة انشاء مجال ثقة لـ R تصبح محصورة في إيجاد نصف طول ذلك المجال ε_β الذي يحقق العلاقة التالية:

$$P\left(r - \varepsilon_\beta < R < r + \varepsilon_\beta\right) = \beta = 1 - \alpha \qquad (27-2)$$

وبطرح r من أطراف المتراجحة وضربها ب (−1) ثم تقسيمها على σ_r نحصل على الشرط المكافئ التالي:

$$P\left(\frac{-\varepsilon_\beta}{\sigma_r} < \frac{r-R}{\sigma_r} < \frac{\varepsilon_\beta}{\sigma_r}\right) = \beta = 1 - \alpha \qquad (28-2)$$

ولحساب الاحتمال الذي في الطرف الأيسر نميز بين الحالتين التاليتين :

2-3-1: الحالة التي يكون فيها σ_r مقدراً (حالة حجم العينة n صغيراً):

عندئذ نجد أن σ_r يقدر بالعلاقة:

$$\tilde{\sigma}_r = \sqrt{\frac{rq}{n-1}} \approx \sqrt{\frac{rq}{n}} \qquad (29-2)$$

وبالتعويض في العلاقة (2-28) نحصل على ما يلي:

$$P\left(\frac{-\varepsilon_\beta}{\sqrt{\frac{rq}{n}}} < \frac{r-R}{\sqrt{\frac{rq}{n}}} < \frac{\varepsilon_\beta}{\sqrt{\frac{rq}{n}}}\right) = 1-\alpha \qquad (30-2)$$

وبمقارنة ذلك مع المتوسط نجد أن المتحول $\frac{r-R}{\sqrt{\frac{rq}{n}}}$ خاضع لتوزيع (ستودينت) ذي (n-1) درجة حرية . وذلك لأن r هي حالة خاصة من المتوسط عندما $\left(x_i = 0 \text{ أو } 1\right)$ واعتماداً على ذلك يمكننا أن نطبق العلاقة (2-11) مباشرة على r فنحصل على نصف الطول ε_β ، وبعد الاستفادة من (2-29) نجد أن:

$$\varepsilon_\beta = t_{\left(1-\frac{\alpha}{2},n-1\right)} * \sqrt{\frac{rq}{n}} \qquad (31-2)$$

وبالتعويض في (2-30) نحصل على ما يلي:

$$P\left(-t_{\left(1-\frac{\alpha}{2}\right)} < \frac{r-R}{\sqrt{\frac{rq}{n}}} < t_{\left(1-\frac{\alpha}{2}\right)}\right) = 1-\alpha \qquad (32-2)$$

وبالتعويض في (2-27) نحصل على مجال الثقة للنسبة R في المجتمع كما يلي:

$$P\left(r - t_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{rq}{n}} < R < r - t_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{rq}{n}}\right) = 1-\alpha \qquad (33-2)$$

وهو مجال الثقة المطلوب .

2-3-2: الحالة التي يكون فيها σ_r معلوماً (حالة حجم العينة n كبيراً)

لمعالجة هذه الحالة نلاحظ أن المتحول $\frac{r-R}{\sigma_r}$ في العلاقة (2-28) وكحالة خاصة من المتوسط، يخضع للتوزيع الطبيعي المعياري N(0 , 1) . وما دامت r حالة خاصة من $\bar{x}$ يمكننا أن نطبق العلاقة (2-19) مباشرة حيث نجد أن:

$$\varepsilon_\beta = z_{1-\frac{\alpha}{2}} * \sqrt{\frac{rq}{n}} \qquad (34-2)$$

وبالتعويض في العلاقة (2-30) نجد أن:

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{r-R}{\sqrt{\frac{rq}{n}}} < z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \qquad (35-2)$$

ومن ثم فإننا سنحصل على مجال الثقة التالي:

$$P\left(r - z_{1-\frac{\alpha}{2}} * \sqrt{\frac{rq}{n}} < R < r + z_{1-\frac{\alpha}{2}} * \sqrt{\frac{rq}{n}}\right) = 1-\alpha \qquad (36-2)$$

وهو المجال المطلوب لـ R .

- **مثال (2-3):** سألنا عشرين طالباً عما إذا كانوا قد نجحوا في الامتحان أم لا . فأجاب (12) منهم بنعم . فما هو تقدير نسبة النجاح وما مجال الثقة الذي يحتوي النسبة الحقيقية للنجاح R باحتمال قدره 0,95 ؟

الحل: إن التقدير الأولي لنسبة النجاح هو:

$$\tilde{R} = r = \frac{12}{20} = 0,60$$

وبما أن حجم العينة $n = 20$ صغير, فإن مسألة إيجاد مجال ثقة لـ R تعتمد على توزيع (ستودينت) . لذلك نحسب نصف طول مجال الثقة من العلاقة :

$$\varepsilon_\beta = t_{1-\frac{\alpha}{2}} * \sqrt{\frac{rq}{n}}$$

ومن جدول (ستودينت) نجد أن:

$$t_{\left(1-\frac{\alpha}{2}, n-1\right)} = t_{(0,975,19)} = 2,09$$

وبذلك يكون :

$$\varepsilon_\beta = 2,09 * \sqrt{\frac{(0,60)(0,40)}{20}} = 0,22895$$

وبذلك نجد أن مجال الثقة المطلوب هو:

$$P(0,60 - 0,22895 < R < 0,60 + 0,22895) = 0,95$$
$$P(0,371 < R < 0,829) = 0,95$$

وهو مجال عريض للنسبة R يقتضي زيادة حجم العينة n .

- **مثال (2-4):** في دراسة على المصابيح الكهربائية أخذنا عينة بحجم $n = 100$ مصباحاً فوجدنا أن (7) منها غير صالحة للاستعمال . أوجد تقدير نسبة العطب في الإنتاج ثم أوجد مجال ثقة لها باحتمال قدره 0,90؟ .

الحل: إن تقدير نسبة العطب هو:

$$\tilde{R} = r = \frac{7}{100} = 0,07$$

وبما أن حجم العينة كبير, فإن مسألة إيجاد مجال ثقة لـ R تعتمد على التوزيع الطبيعي المعياري . لذلك نحسب نصف طول ذلك المجال من العلاقة :

$$\varepsilon_\beta = z_{1-\frac{\alpha}{2}} * \sqrt{\frac{rq}{n}}$$

ومن جدول $\phi(z)$ الطبيعي المعياري نجد أن:

$$z_{1-\frac{\alpha}{2}} = z_{0,95} = 1,65$$

وبذلك يكون

$$\varepsilon_\beta = 1,65 * \sqrt{\frac{(0,07)(0,93)}{100}} = 0,0255$$

وإن مجال الثقة المطلوب هو:

$$P(0,07 - 0,0255 < R < 0,07 + 0,0255) = 0,90$$

أي أن المجال المطلوب لنسبة العطب هو [0,0445 , 0,0955] .

2-4: إنشاء مجال ثقة لإجمالي مجتمع طبيعي:

نعلم أن إجمالي المجتمع Y يساوي: $Y = N\ \mu$ وأن تقديره يعطى بالعلاقة التالية :

$$\tilde{Y} = N\,\tilde{\mu} = N\,\bar{x} \qquad (37-2)$$

وإن تقدير تباين $\tilde{Y}$ يعطى بالعلاقة:

$$\tilde{\sigma}_{\tilde{Y}}^2 = N^2 * \tilde{\sigma}_{\tilde{Y}}^2 = N^2 * \frac{s^2}{n} \qquad (38-2)$$

ولإنشاء مجال للثقة يحوي Y باحتمال ثقة يساوي β سنبحث عن ε_β الذي يحقق العلاقة :

$$P\left(\tilde{Y} - \varepsilon_\beta < Y < \tilde{Y} + \varepsilon_\beta\right) = \beta \qquad (39-2)$$

أو العلاقة:

$$P\left(\frac{-\varepsilon_\beta}{\sigma_{\tilde{Y}}} < \frac{\tilde{Y} - Y}{\sigma_{\tilde{Y}}} < \frac{\varepsilon_\beta}{\sigma_{\tilde{Y}}}\right) = \beta \qquad (40-2)$$

وبمعالجة هذه المسألة بطريقة مشابهة لما فعلناه في حالة المتوسط نجد أن ε_β يحسب بإحدى طريقتين متأثرتين بحجم العينة n , وهنا نميز الحالتين التاليتين :

2-4-1: الحالة التي يكون فيها $\sigma_{\tilde{Y}}$ مقدراً (أو حالة حجم العينة n صغيراً)

في هذه الحالة يكون $\frac{\tilde{Y}-Y}{\tilde{\sigma}_{\tilde{Y}}^2}$ خاضعاً لتوزيع (ستودينت) $T_{(n-1)}$ ذي $(n-1)$ درجة حرية . وقياساً على العلاقة (2-11) فإننا نجد أن ε_β يحسب من العلاقة :

$$\varepsilon_\beta = t_{\left(1-\frac{\alpha}{2},n-1\right)} * N\frac{s}{\sqrt{n}} \qquad (41-2)$$

وبالتعويض في (2-40) نجد أن:

$$P\left(-t_{\left(1-\frac{\alpha}{2}\right)} < \frac{\tilde{Y} - Y}{N\ s/\sqrt{n}} < t_{\left(1-\frac{\alpha}{2}\right)}\right) = \beta = 1 - a \qquad (42-2)$$

ويكون مجال الثقة المطلوب هو:

$$P\left(\tilde{Y} - t_{1-\frac{\alpha}{2}} * N * \frac{s}{\sqrt{n}} < Y < \tilde{Y} + t_{1-\frac{\alpha}{2}} * N * \frac{s}{\sqrt{n}}\right) = \beta = 1 - \alpha \qquad (43-2)$$

2-4-2: الحالة التي يكون فيها $\sigma_{\tilde{Y}}$ معلوماً (حالة حجم العينة n كبيراً)

في هذه الحالة يكون المتحول $\frac{\tilde{Y}-Y}{\sigma_{\tilde{Y}}}$ خاضعاً للتوزيع الطبيعي المعياري, وقياساً على العلاقة (2-19) يكون لدينا

$$\varepsilon_\beta = z_{1-\frac{\alpha}{2}} * N * \frac{s}{\sqrt{n}} \qquad (44-2)$$

وبالتعويض في (2-40) نحصل على أن:

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\tilde{Y}-Y}{N\ s/\sqrt{n}} < z_{1-\frac{\alpha}{2}}\right) = \beta = 1-\alpha \qquad (45-2)$$

ويكون مجال الثقة لهذه الحالة مساوياً:

$$P\left(\tilde{Y} - z_{1-\frac{\alpha}{2}} * N * \frac{s}{\sqrt{n}} < Y < \tilde{Y} + z_{1-\frac{\alpha}{2}} * N * \frac{s}{\sqrt{n}}\right) = 1-\alpha \qquad (46-2)$$

2-5: إنشاء مجال ثقة للفرق بين متوسطي مجتمعين طبيعيين :

وتعالج هذه المسألة بطريقة مشابهة تماماً لتلك التي اوردناها في حالة متوسط واحد ونصيغها على الشكل التالي:

لنفترض أنه لدينا مجتمعان A_1 وA_2 يخضعان للتوزيعين الطبيعيين $N(\mu_1, \sigma_1^2)$ و$N(\mu_2, \sigma_2^2)$ على الترتيب .

كما نفترض أننا سحبنا منهما عينتين بحجمين n_1 وn_2 وكان متوسطا هاتين العينتين $\bar{x}_1$ و$\bar{x}_2$ وتبايناهما المصححان s_1^2 وs_2^2 على الترتيب, والمطلوب إنشاء مجال ثقة للفرق $(\mu_1-\mu_2)$.

إننا نعلم أن تقدير الفرق $(\mu_1-\mu_2)$ هو:

$$\tilde{\Delta} = (\widetilde{\mu_1-\mu_2}) = \bar{x}_1 - \bar{x}_2 \qquad (47-2)$$

وإن تباين هذا التقدير يساوي:

$$\sigma_{\tilde{\Delta}}^2 = var(\bar{x}_1-\bar{x}_2) = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \qquad (48-2)$$

ولإنشاء مجال ثقة لـ $(\mu_1-\mu_2)$ نحتاج إلى إيجاد ε_β الذي يحقق العلاقة :

$$P\left((\bar{x}_1-\bar{x}_2) - \varepsilon_\beta < (\mu_1-\mu_2) < (\bar{x}_1-\bar{x}_2) + \varepsilon_\beta\right) = \beta \qquad (49-2)$$

التي يمكن كتابتها على الشكل:

$$P\left(\frac{-\varepsilon_\beta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < \frac{+\varepsilon_\beta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right) = \beta \qquad (50-2)$$

ولمتابعة المعالجة سنميز بين الحالتين التاليتين :

2-5-1: الحالة التي يكون فيها $\sigma_{\tilde{\Delta}}$ مقدراً (حالة حجم العينة الكلية $n = n_1 + n_2$ صغيراً):

في هذه الحالة نعلم أن تقدير σ_{Δ}^2 يساوي:

$$\tilde{\sigma}_{\Delta}^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \qquad (51-2)$$

وعندئذ فإن العلاقة (2-50) تكتب على الشكل التالي :

$$P\left(\frac{-\varepsilon_\beta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} < \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} < \frac{\varepsilon_\beta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right) = \beta \qquad (52-2)$$

ولحساب ε_β يقتضي الأمر أن يكون التوزيع الاحتمالي للمقدار الأوسط معروفاً .

ولكن التوزيع الاحتمالي للمتحول $\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ مازال غير معروف وهو يشكل مسألة مازالت قائمة وتسمى بمسألة (برينز– فيشر), وإن مسألة إيجاد مجال ثقة لـ $(\mu_1 - \mu_2)$ ضمن الشروط العامة مازالت غير محلولة, إلا في بعض الحالات الخاصة ومع بعض الشروط الإضافية.

فمثلاً يمكننا أن نجد مجال ثقة لـ $(\mu_1 - \mu_2)$ لبعض المجتمعات التي تتصف بشروط إضافية وهي أن يكون تباينا المجتمعين A_1 وA_2 متساويين, أي عندما يكون لدينا:

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 \qquad (53-2)$$

لأن مسألة إنشاء مجال ثقة للفرق $(\mu_1 - \mu_2)$ تصبح قابلة للحل وتعالج كما يلي:

بما انه لدينا $\sigma_1^2 = \sigma_2^2 = \sigma^2$, وأن تبايني العينين s_1^2 وs_2^2 هما تقديران لكل من σ_1^2 وσ_2^2 على الترتيب, ومن ثم لـ σ^2 , فإنه يمكننا أن نقدر σ^2 بأي من التباينين s_1^2 وs_2^2 أو بمتوسطيهما , ولكننا ولأسباب فنية (كعدم التحيز والكفاية وضرورة التوصل إلى متحول ذي توزيع معروف) فإننا نقدر σ^2 بوساطة العلاقة التالية :

$$\tilde{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = s^2 \qquad (54-2)$$

وهو تقدير غير متحيز لـ σ^2 .

وضمن هذه الشروط فإن التباين $\sigma_{\tilde{\Delta}}^2$ المعرف في (2-48) يصبح مساوياً لـ :

$$\sigma_{\tilde{\Delta}}^2 = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right) \qquad (55-2)$$

وأن تقديره غير المتحيز يعطى بالعلاقة التالية :

$$\tilde{\sigma}_{\tilde{\Delta}}^2 = s^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \qquad (56-2)$$

وبالعودة إلى العلاقة (2−50) نجد أنها تأخذ الشكل التالي:

$$\left(\frac{-\varepsilon_\beta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}} < \frac{\varepsilon_\beta}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \right) = \beta \qquad (57-2)$$

وبملاحظة أن المقدار الأوسط خاضع لتوزيع (ستودينت) ذي (n_1+n_2-2) درجة حرية (راجع النظريات في كتاب نظرية الاحتمالات) . وقياساً على العلاقة (2−11) نجد أن نصف طول المجال المطلوب يعطى بالعلاقة:

$$\varepsilon_\beta = t_{\left(1-\frac{\alpha}{2},n_1+n_2-2\right)} * s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \qquad (58-2)$$

وبالتعويض في (2−57) نحصل على العلاقة:

$$P\left(-t_{\left(1-\frac{\alpha}{2},n_1+n_2-2\right)} < \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < t_{\left(1-\frac{\alpha}{2},n_1+n_2-2\right)} \right) = 1-\alpha \qquad (59-2)$$

وبذلك نحصل على مجال الثقة التالي:

$$P\left((\bar{x}_1-\bar{x}_2) - t_{1-\frac{\alpha}{2}} * s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + t_{1-\frac{\alpha}{2}} * s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \right) = 1-\alpha \qquad (60-2)$$

وهو المجال المطلوب والمشروط بأن يكون $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

2−5−2: الحالة التي يكون فيها $\sigma_{\bar{\Delta}}$ معلوماً (حالة حجم العينة الكلية $n = n_1 + n_2$ كبيراً):

في هذه الحالة لا نشترط أن يتساوى التباينات σ_1^2 و σ_2^2 , ولكننا سنعتبر المقدار $\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}$ قيمة تقريبية للمقدار $\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$ الوارد في العلاقة (2−50) وليس متحولاً عشوائياً . واعتماداً على ذلك فإن المقدار .

$$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \approx \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \quad (n \text{ كبيرة}) \qquad (61-2)$$

ويكون خاضعاً للتوزيع الطبيعي المعياري $N(0\,,1)$.

وقياساً على العلاقة (2−19) نجد أن ε_β يحسب من العلاقة التالية :

$$\varepsilon_\beta = z_{1-\frac{\alpha}{2}} * \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \qquad (62-2)$$

وبالتعويض في (2–50) نحصل على العلاقة التالية:

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \qquad (63-2)$$

ومن ثم فإننا سنحصل على مجال الثقة التالي :

$$P\left((\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right) = 1 - \alpha \quad (64-2)$$

2-5-3: حالة خاصة: سحب عينتين من مجتمع واحد واختبار معنوية الفرق $(\bar{x}_1 - \bar{x}_2)$.

عندما نسحب العينتين السابقتين من مجتمع واحد فإنه يكون لدينا مايلي:

$$\left.\begin{array}{l}\mu_1 = \mu_2 = \mu \implies \mu_1 - \mu_2 = 0 \\ \sigma_1^2 = \sigma_1^2 = \sigma^2\end{array}\right\} \qquad (65-2)$$

وعندئذ فإننا نقدر التباين σ^2 بالعلاقة :

$$\tilde{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = s^2 \qquad (66-2)$$

وهنا نميز بين الحالتين السابقتين :

- **عندما تكون $n = n_1 + n_2$ صغيراً :**

في هذه الحالة نجد أن العلاقة (2–59) تأخذ الشكل التالي:

$$P\left(-t_{\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)} < \frac{\bar{x}_1 - \bar{x}_2 - 0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t_{\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)}\right) = 1 - \alpha$$

وهو يعطينا احتمال أن يكون :

$$\left|\frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right| < t_{\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)} \qquad (67-2)$$

يساوي $(1 - \alpha)$.

وبذلك تنقلب مسألة إنشاء مجال الثقة إلى مسألة اختبار معنوية أو جوهرية الفرق $(\bar{x}_2 - \bar{x}_1)$. ولإجراء هذا الاختبار بمستوى دلالة α نقوم بحساب المقدار t حيث :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (68-2)$$

ثم نقارنه بـ $t_{\left(1-\frac{a}{2}, n_1+n_2-2\right)}$ الجدولية فإذا كان :

$$|t| < t_{\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)} \qquad (69-2)$$

فإننا نقول عن الفرق $(\bar{x}_1 - \bar{x}_2)$ بين متوسطي العينتين أنه فرق غير جوهري وأنه عائد لأسباب عرضية عشوائية .

أما إذا كان $|t| \geq t_{\left(1-\frac{a}{2}\right)}$ فإننا نحكم بأن الفرق $(\bar{x}_1 - \bar{x}_2)$ جوهري وعائد لأسباب موضوعية يمكن البحث عنها وتحديدها .

- **عندما تكون $n = n_1 + n_2$ كبيراً :**

إن هذه الحالة لا تختلف كثيراً عن الحالة السابقة(n صغيراً), إلا في أن التوزيع الاحتمالي المستخدم للمقارنة سيكون التوزيع الطبيعي المعياري $N(0\,,1)$ بدلاً من توزيع (ستودينت), فلاختبار معنوية الفرق $(\bar{x}_1 - \bar{x}_2)$ بمستوى دلالة α ، نحسب المقدار Z حيث :

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (70-2)$$

ثم نقارنه بـ $Z_{\left(1-\frac{\alpha}{2}\right)}$ الجدولية فإذا كان :

$$|Z| < Z_{\left(1-\frac{\alpha}{2}\right)} \qquad (71-2)$$

فإننا نقول عن الفرق $(\bar{x}_1 - \bar{x}_2)$ إنه فرق غير جوهري وإنه عائد لأسباب عرضية عشوائية .

أما إذا كان $|Z| \geq Z_{\left(1-\frac{\alpha}{2}\right)}$ فإننا نحكم بأن الفرق $(\bar{x}_1 - \bar{x}_2)$ جوهري وعائد لأسباب موضوعية يمكن البحث عنها وتحديدها .

- **مثال (2−5)**: لنفترض أننا سحبنا عينتين بحجمين $n_1 = 22\,, n_2 = 20$ من مدرستين لدراسة الفرق بين متوسطي علامة الطالب في مقرر الرياضيات في هاتين المدرستين . فوجدنا أن متوسطي هاتين العينتين وتباينهما كانا يساويان :

$$\bar{x}_1 = 60 \qquad \bar{x}_2 = 70$$
$$s_1^2 = 100 \qquad s_2^2 = 200$$

والمطلوب إيجاد مجال ثقة وباحتمال قدره 0.95 للفرق بين متوسطي العلامة في هذين المجتمعين .

الحل: إن هذه المسألة تندرج تحت الفقرة (2−5−2)لأن حجم العينة الكلية $n = 22 + 20$ يعتبر كبيراً نسبياً. لذلك فإننا نجد أن نصف طول مجال الثقة ذي الاحتمال 0.95 يساوي (انظر (2−62)) .

$$\varepsilon_\beta = Z_{\left(1-\frac{\alpha}{2}\right)} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1{,}96 * \sqrt{\frac{100}{20} + \frac{200}{22}} = 7{,}36$$

ومجال الثقة المطلوب هو :

$$P[(60-70) - 7{,}36 < \mu_1 - \mu_2 < (60-70) + 7{,}36] = 0{,}95$$

أي أن المجال المطلوب هو $[-17{,}36\ ,\ -2{,}64]$.وهو يشير إلى أن $\mu_1 < \mu_2$.

- **مثال (6-2):** لنفترض أننا سحبنا عينتين بحجمين $n_1 = 10 , n_2 = 15$ من الخراف ومن قطيع واحد فوجدنا متوسط وزن خراف العينة الأولى $\bar{x}_1 = 30$ ومتوسط وزن خراف العينة الثانية $\bar{x}_2 = 33$ وإن تباينهما يساويان $s_1^2 = 300$ و $s_2^2 = 400$.

فهل هناك فرق جوهري وبمستوى دلالة $\alpha = 0.05$ بين متوسطي العينتين من حيث الوزن ؟

الحل: إن هذه المسألة تندرج تحت الفقرة (2−5−3) ولذلك نقوم بحساب s^2 المدمج , حيث نجد أن:

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{9*300+14*400}{10+15-2} = 360,87$$

وبما أن حجم العينة الكلية صغير نسبياً $10+15 < 30$,فإننا نطبق (2−68) ونحسب t فنجد أن:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{30-33}{\sqrt{360,87}\sqrt{\frac{1}{10}+\frac{1}{15}}} = -0,38676$$

ثم نقوم بأيجاد $t_{1-\frac{\alpha}{2}}$ المقابلة n_1+n_2-2 درجة حرية من جدول توزيع (ستودينت) فنجد أن:

$$t_{(0,975,23)} = 2,07$$

وبمقارنة t مع $t_{1-\frac{\alpha}{2}}$ نجد أن:

$$0,38676 = |\mathrm{t}| < \mathrm{t}_{1-\frac{\alpha}{2}} = 2,07$$

إذن نحكم بأن الفرق بين متوسطي العينتين هو فرق غير جوهري وأنه عائد لأسباب عرضية وعشوائية .

2− 6: إنشاء مجال ثقة للفرق بين نسبتين $\mathbf{R_1}$ و $\mathbf{R_2}$ لخاصة واحدة في مجتمعين طبيعيين :

لقد أشرنا سابقاً إلى أنه عندما نسحب عينتين بحجم n_1 و n_2 من مجتمعين فإن النسبتين r_1 و r_2 هما تقديران لـ R_1 و R_2 على الترتيب . ولقد رأينا أنه يمكننا اعتبار النسبة r في العينة حالة خاصة لمتوسط متحول عشوائي X يأخذ إحدى القيمتين: 1 أو 0 .

واعتماداً على ذلك وعلى العلاقات (2−23) و(2−24) و(2−25) وكذلك العلاقات (2−46) و(2−48) و(2−49) و(2−50) نجد أن المجال المطلوب هو المجال الذي نصف طوله ε_β ويحقق العلاقة التالية:

$$P\left[(r_1-r_2)-\varepsilon_\beta < (R_1-R_2) < (r_1-r_2)+\varepsilon_\beta\right] = \beta \qquad (72-2)$$

وبملاحظة أن تباين الفرق (r_1-r_2) يساوي :

$$\sigma_{\bar{\Delta}}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

فإننا نجد أن العلاقة السابقة تكافئ أن يكون:

$$P\left[\frac{-\varepsilon_\beta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < \frac{(r_1-r_2)-(R_1-R_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < \frac{\varepsilon_\beta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right] = \beta \qquad (73-2)$$

وبطريقة مشابهة لما فعلناه عند دراسة مجال الثقة للفرق بين متوسطين سنميز بين الحالتين التاليتين :

2-6-1: عندما يكون $n = n_1 + n_2$ صغيراً

فإن المسألة مازالت غير قابلة للحل إلا إذا كان لدينا

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

وهذا ما يكافئ في حالتنا هذه أن تحقق النسب في المجتمعين الشرط التالي :

$$R_1\, Q_1 = R_2\, Q_2 = \sigma^2 \qquad (74-2)$$

ولنفترض أن هذا الشرط محقق في المجتمعين المدروسين , وعندئذ نقوم بتقدير σ^2 بوساطة علاقة ناتجة عن (2-54) وهي العلاقة:

$$\tilde{\sigma}^2 = s^2 = \frac{n_1 r_1 q_1 + n_2 r_2 q_2}{n_1 + n_2 - 2} = \qquad (75-2)$$

وعندئذ نحصل على:

$$\tilde{\sigma}_{\tilde{\Delta}}^2 = \tilde{\sigma} * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad (76-2)$$

وبذلك تأخذ العلاقة (2-73) الشكل التالي:

$$P\left(\frac{-\varepsilon_\beta}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < \frac{(r_1 - r_2) - (R_1 - R_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < \frac{\varepsilon_\beta}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right) = \beta \qquad (77-2)$$

وبما ان المتحول الأوسط يخضع لتوزيع (ستودينت) ذي $n_1 + n_2 - 2$ درجة حرية . فإننا اعتماداً على (2-11) أو (2-58) نجد أن ε_β يساوي:

$$\varepsilon_\beta = t_{\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)} * s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad (78-2)$$

حيث أن S تحسب من العلاقة (2-75) أما $t_{1-\frac{\alpha}{2}}$ فتؤخذ من جدول (ستودينت) .

وبالتعويض في (2-73) نحصل على أن:

$$P\left(-t_{\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)} < \frac{(r_1 - r_2) - (R_1 - R_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t_{\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)}\right) = \beta \qquad (79-2)$$

وبذلك نحصل على مجال الثقة التالي:

$$\boldsymbol{P\left((r_1 - r_2) - t_{\left(1-\frac{\alpha}{2}\right)} * s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < R_1 - R_2 < (r_1 - r_2) + t_{\left(1-\frac{\alpha}{2}\right)} * s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = \beta \quad (80-2)}$$

وهو المجال المطلوب والمشروط بأن يكون $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

2-6-2: عندما يكون $n = n_1 + n_2$ كبيراً :

في هذه الحالة لا نشترط أن يتساوى التباينان σ_1^2 و σ_2^2 , ولكننا سنعتبر المقدار r_1q_1 قيمة تقريبية لـ σ_1^2 وr_2q_2 قيمة تقريبية لـ σ_2^2 . وبذلك فإن العلاقة (2-73) تأخذ الشكل التالي:

$$P\left(\frac{-\varepsilon_\beta}{\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}} < \frac{(r_1-r_2)-(R_1-R_2)}{\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}} < \frac{\varepsilon_\beta}{\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}}\right) = \beta \qquad (81-2)$$

والمقدار الأوسط يكون خاضعاً للتوزيع الطبيعي المعياري $N(0\,,1)$, وقياساً على العلاقة (2-19) نجد أن ε_β تحسب من العلاقة :

$$\varepsilon_\beta = Z_{1-\frac{\alpha}{2}} * \sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}} \qquad (82-2)$$

وبالتعويض في (2-81) نحصل على :

$$P\left(-Z_{1-\frac{\alpha}{2}} < \frac{(r_1-r_2)-(R_1-R_2)}{\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}} < Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \qquad (83-2)$$

ومنها نجد أن مجال الثقة المطلوب هو:

$$\boldsymbol{P}\left[(\boldsymbol{r_1}-\boldsymbol{r_2}) - \boldsymbol{Z}_{\left(1-\frac{\alpha}{2}\right)} * \sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}} < \boldsymbol{R_1}-\boldsymbol{R_2} < (\boldsymbol{r_1}-\boldsymbol{r_2}) + \boldsymbol{Z}_{\left(1-\frac{\alpha}{2}\right)} * \sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}\right] = \boldsymbol{1}-\alpha \quad (\boldsymbol{84-2})$$

3-6-2: حالة خاصة: اختبار معنوية الفرق $(r_1 - r_2)$:

عندما نسحب العينين السابقتين من مجتمع واحد فإنه يكون لدينا:

$$\left.\begin{matrix} R_1 = R_2 = R \\ \sigma_1^2 = \sigma_2^2 = \sigma^2 \end{matrix}\right\}$$

وعندئذ نقدر التباين الموحد σ^2 بالعلاقة :

$$\tilde{\sigma}^2 = s^2 = \frac{n_1r_1q_1 + n_2r_2q_2}{n_1+n_2-2} = \qquad (85-2)$$

ولاختبار معنوية الفرق $(r_1 - r_2)$ فإننا نقوم بحساب المقدار

$$\frac{r_1-r_2}{\boldsymbol{s}\sqrt{\frac{\boldsymbol{1}}{\boldsymbol{n_1}}+\frac{\boldsymbol{1}}{\boldsymbol{n_2}}}} \qquad (86-2)$$

ولإجراء المقارنة نميز بين الحالتين التاليتين :

- **الحالة التي يكون فيها حجم العينتين $n = n_1 + n_2$ صغيراً :**

نقوم بمقارنة المقدار (2-86) مع قيمة متحول (ستودينت) $t_{\left(1-\frac{\alpha}{2},n_1+n_2-2\right)}$, فإذا كانت قيمة المقدار :

$$\left|\frac{r_1 - r_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right| < t_{\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)} \qquad (87-2)$$

فإننا نقول أن الفرق $(r_1 - r_2)$ غير جوهري وعائداً لأسباب عرضية أو عشوائية .

أما إذا كان العكس فنقول عن الفرق $(r_1 - r_2)$ أنه فرق جوهري وعائد لأسباب موضوعية يمكن البحث عنها وتحديدها .

- **الحالة التي يكون فيها حجم العينتين $n = n_1 + n_2$ كبيراً :**

نقوم بمقارنة المقدار (2–86) مع قيمة متحول التوزيع الطبيعي المعياري $Z_{\left(1-\frac{\alpha}{2}\right)}$, فإذا كانت قيمة المقدار :

$$\left|\frac{r_1 - r_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right| < Z_{\left(1-\frac{\alpha}{2}\right)} \qquad (88-2)$$

فإننا نقول أن الفرق $(r_1 - r_2)$ غير جوهري وعائداً لأسباب عرضية أو عشوائية .

أما إذا كان العكس فنقول عن الفرق $(r_1 - r_2)$ أنه فرق جوهري وعائد لأسباب موضوعية يمكن البحث عنها وتحديدها .

- **ملاحظة:** يمكننا هنا تقدير المقدار s بوساطة العلاقة $S\sqrt{r * q}$ حيث أن $r = \frac{n_1 r_1 + n_2 r_2}{n_1 + n_2}$ وأن $q = 1 - r$
- **مثال (2–7):** لدراسة نسبتي المدخنين في مجتمعين يتميزان بأن فيهما $\sigma_1^2 = \sigma_2^2$, سحبنا منهما عينتين بحجمين $n_1 = 10$ و$n_2 = 20$ فكان عدد المدخنين في العينة الأولى 4 وفي الثانية 5 أشخاص . فأوجد مجال الثقة للفرق بين نسبتي المدخنين في المجتمعين باحتمال قدره 0.95 .

الحل: نلاحظ أولاً أن مجموع حجمي العينتين 10+20 صغير نسبياً . وبما أن تبايني المجتمعين متساويان . إذن فسوف يمكننا تطبيق العلاقة (2–78) لحساب نصف طول مجال الثقة المطلوب . ولهذا نبحث عن قيمة $t_{1-\frac{\alpha}{2}}$ المقابلة لـ (2–10+20) درجة حرية فنجد أن:

$$t_{(0{,}975\ ,28\ )} = 2{,}05$$

ونحسب s من العلاقة (2–75) فنجد أن :

$$s \quad = \sqrt{\frac{10\,\frac{4}{10} * \frac{6}{10} + 20 * \frac{5}{20} - \frac{15}{20}}{10 + 20 - 2}} = 0{,}449$$

وبذلك نجد أن :

$$\varepsilon_\beta = (2{,}05)(2{,}449)\sqrt{\frac{1}{10} + \frac{1}{20}} = 0{,}3565$$

وبذلك نجد أن:

$$P[(0{,}4 - 0{,}25) - 0{,}3565 < R_1 - R_2 < (0{,}4 - 0{,}25) + 0{,}3565] = 0{,}95$$

أي أن المجال المطلوب هو $\left[0{,}506 \quad و \quad 0{,}206\right]$.

وهو يشير إلى فرق محسوس أوجوهري بين نسبتي المدخنين في المجتمعين.

- **مثال (2-8):** عند دراسة نسبة العطب في إنتاج معملين ينتجان البضاعة نفسها أخذنا عينة من منتوجات المعمل الأول بحجم $n_1 = 30$ قطعة وعينة من منتوجات المعمل الثاني بحجم $n_2 = 40$, فوجدنا أن نسبتي العطب كانتا $r_1 = 0{,}07$ و$r_2 = 0{,}09$. فأوجد مجال الثقة للفرق بين نسبتي العطب في الإنتاج في كلا المعملين وذلك باحتمال قدره 0.95 .

الحل: نلاحظ ان $n_1 + n_2 = 70$ كبير نسبياً . لذلك نطبق العلاقة $(2-82)$ لحساب ε_β , ولهذا نبحث عن $Z_{1-\frac{\alpha}{2}}$ فنجد أن :

$$Z_{1-\frac{\alpha}{2}} = Z_{0{,}95} = 1{,}65$$

$$\tilde{\sigma}_{\tilde{\Delta}} = \sqrt{\frac{r_1 q_1}{n_1} + \frac{r_2 q_2}{n_2}} = \sqrt{\frac{(0{,}07)(0{,}93)}{30} + \frac{(0{,}09)(0{,}91)}{40}} = 0{,}0649$$

$$\varepsilon_\beta = (1{,}65)(0{,}0649) = 0{,}1071$$

وبذلك نجد أن:

$$P[(0{,}07 - 0{,}09) - 0{,}1071 < R_1 - R_2 < (0{,}07 - 0{,}09) + 0{,}1071] = 0{,}90$$

والمجال المطلوب هو : $[-0{,}1271 \quad , \quad 0{,}0871]$

2-7: إنشاء مجال ثقة لتباين مجتمع طبيعي:

لقد كنا عند إنشاء أي مجال ثقة للمؤشرات السابقة، نقوم بتحديد عنصرين له: مركزه ونصف طوله . وكنا نعتبر تقدير المؤشر المجهول مركزاً للمجال أما نصف طوله فكنا نحسبه اعتماداً على التوزيع الاحتمالي الذي نتوصل إليه .

أما عند إنشاء مجال ثقة لتباين المجتمع σ^2 فإننا سنتبع أسلوباً آخر وهو تحديد الحدين الأيسر والأيمن للمجال المطلوب, بحيث يحتوي ذلك المجال على التباين σ^2 باحتمال معين قدره β .

إن سبب لجوئنا إلى هذا الأسلوب ناتج عن كوننا سنتعامل مع التوزيع الاحتمالي χ^2 الذي يتصف بعدم التناظر . وبناءً على ما تقدم يصبح المطلوب هو البحث عن حدي المجال $\left]D_1 \text{ و } D_2\right[$ حيث أن $D_1 < D_2$ وأن $D_1 > 0$, وبحيث يتحقق لدينا الشرط التالي :

$$P(D_1 < \sigma^2 < D_2) = \beta \qquad (89-2)$$

لنأخذ مقلوب أطراف المتراجحة ثم نضربها بالمقدار $(n-1)s^2$، حيث أن s^2 هو تباين العينة وn حجمها، فنحصل على الشرط المكافئ التالي:

$$P\left(\frac{(n-1)s^2}{D_2} < \frac{(n-1)s^2}{\sigma^2} < \frac{(n-1)s^2}{D_1}\right) = \beta \qquad (90-2)$$

وبما أن المتحول $\frac{(n-1)S^2}{\sigma^2}$ خاضع للتوزيع $\chi^2_{(n-1)}$ (انظر النظريات في نظرية الاحتمالات) . إذن فإن حساب الاحتمال الذي في الطرف الأيسر يعتمد على التوزيع $\chi^2_{(n-1)}$. ولتسهيل العمليات الرياضية نرمز لـ $\frac{(n-1)S^2}{\sigma^2}$ بالرمز χ , وهكذا فإنه يمكننا أن نكتب (2−90) على الشكل التالي :

$$P\left[\frac{(n-1)s^2}{D_2} < \chi < \frac{(n-1)s^2}{D_1}\right) = \beta \qquad (91-2)$$

$$P\left[\chi < \frac{(n-1)s^2}{D_1}\right] - P\left[\chi \le \frac{(n-1)s^2}{D_2}\right] = \beta = 1-\alpha \qquad (92-2)$$

وبمقارنة هذه العلاقة مع تابع التوزيع التجميعي لـ χ , الذي يعطينا قيم χ_p الحدية المقابلة للاحتمال اليساري P , والذي يكون له الصيغة التالية:

$$P(\chi < \chi_p) = \int_0^{\chi_p} \chi^2_n(\chi)dx = P \qquad (93-2)$$

فإنه يمكننا أن نستنتج ونضع أن:

$$P\left(\chi < \frac{(n-1)s^2}{D_1}\right) = P_1 \quad ; \quad P\left(\chi \le \frac{(n-1)s^2}{D_2}\right) = P_2 \qquad (94-2)$$

وبذلك نحصل على المعادلة التالية:

$$P_1 - P_2 = 1-\alpha \qquad (95-2)$$

وبملاحظة أن المعادلة (2−95) تتضمن مجهولين P_1 وP_2 وهي وحيدة، إذن فسيكون لها لا نهاية من الحلول. فلإيجاد أحد تلك الحلول لابد من تحديد أحد المجهولين ثم حساب الآخر، ولذلك نثبت أحدهما و نترك على جانبي مجال الثقة احتمالين متساويين يساويان $\frac{\alpha}{2}$ أي نضع :

$$P_2 = P\left[\chi \le \frac{(n-1)s^2}{D_2}\right] = \frac{\alpha}{2} \qquad (96-2)$$

وبتعويض ذلك في (2−95) نحصل على :

$$P_1 = P\left[\chi < \frac{(n-1)s^2}{D_1}\right] = 1-\frac{\alpha}{2} \qquad (97-2)$$

ويمكننا حل هاتين المعادلتين وإيجاد حلهما اعتماداً على جداول التوزيع $\chi^2_{(n-1)}(x)$ ، فمن مقارنتهما بالعلاقة (2−94) نجد أن حليهما يساويان على التوالي :

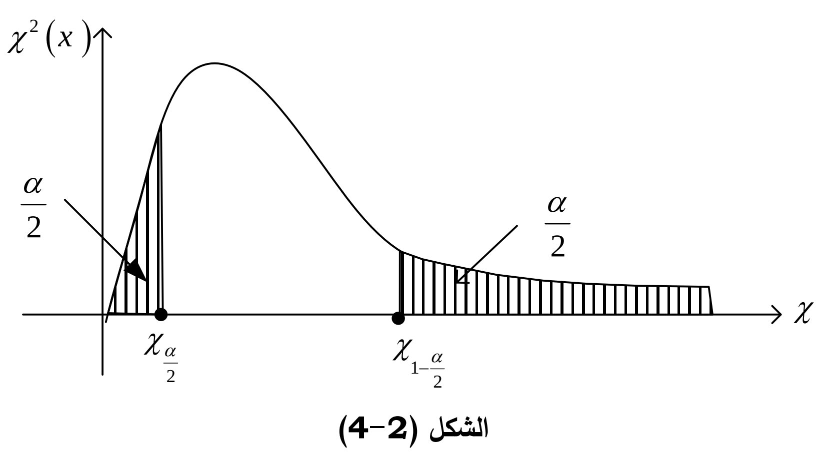
$$\frac{(n-1)s^2}{D_2} = \chi_{\frac{\alpha}{2}} \qquad (98-2)$$

$$\frac{(n-1)s^2}{D_1} = \chi_{1-\frac{\alpha}{2}} \qquad (99-2)$$

حيث أن $\chi_{\frac{\alpha}{2}}$ هي القيمة الجدولية للمتحول χ المقابلة لـ $\frac{\alpha}{2}$ ولـ $(n-1)$ درجة حرية .

وأن $\chi_{1-\frac{\alpha}{2}}$ هي القيمة الجدولية للمتحول χ المقابلة لـ $1-\frac{\alpha}{2}$ ولـ $(n-1)$ درجة حرية .

والشكل (2−4) يوضح لنا معنى هاتين القيمتين وموقعهما .



الشكل (2-4)

ومن العلاقتين (2-95) و(2-99) نقوم بحساب D_1 وD_2 فنجد أن:

$$D_2 = \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}} \qquad (100-2)$$

$$D_1 = \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}} \qquad (101-2)$$

وبالتعويض في العلاقة (2-89) نحصل على مجال الثقة التالي :

$$P\left[\frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}}\right] = 1 - a \qquad (102-2)$$

وهو المجال الذي يمكن أن يحتوي على σ^2 باحتمال قدره $1 - \alpha$ ويترك على طرفي المجال احتمالاً قدره $\frac{\alpha}{2}$.

- **مثال(2-9)**: أوجد مجال ثقة لتباين طول الطالب σ^2 علماً بأن تباين العينة التي حجمها $n = 10$ كان يساوي $s^2 = 24{,}28$ وذلك باحتمال ثقة قدره 0.90 .

الحل: نقوم بحساب D_1 وD_2 من العلاقتين (2-100) و(2-101) لذلك نبحث عن قيمتي $\chi_{\frac{\alpha}{2}}$ و$\chi_{1-\frac{\alpha}{2}}$ عند درجة الحرية 9 = 1- 10, فنجد أن:

$$\chi_{\left(1-\frac{\alpha}{2},9\right)} = \chi_{(0{,}95{,}9)} = 16{,}92$$
$$\chi_{\frac{\alpha}{2}} = \chi_{(0{,}05{,}9)} = 3{,}33$$

وبالتعويض في (2-100) و(2-101) نجد أن:

$$D_1 = \frac{9 * 24{,}28}{16{,}92} = 12{,}91$$
$$D_2 = \frac{9 * 24{,}28}{3{,}33} = 65{,}82$$

مجال الثقة المطلوب هو :

$$P(12{,}91) < \sigma^2 < 95{,}82) = 0{,}90$$

8-2: إنشاء مجال ثقة لنسبة تبايني مجتمعين طبيعيين .

لنفترض أنه لدينا مجتمعان طبيعيان، تباينهما σ_1^2 و σ_2^2 مجهولان, ونريد أن ننشئ مجال ثقة للنسبة $\frac{\sigma_1^2}{\sigma_2^2}$,حيث أن مقارنة التباينين σ_1^2 و σ_2^2 تفيدنا كثيراً في العديد من البحوث العلمية, لمعرفة الدقة التقنية لنوعين من الأجهزة، أو لدراسة تجانس المجتمعين قبل اختبار الفرق بين متوسطيهما.

فإذا كانت النسبة $\frac{\sigma_1^2}{\sigma_2^2}$ أو تقديرها $\frac{s_1^2}{s_2^2}$ مساوية أو قريبة من الواحد فإن ذلك يعني أن المجتمعين متشابهان من حيث التباين. أما إذا كانت أصغر من الواحد أو أكبر منه فإن ذلك يقدم لنا دلالة على أن تباينيهما مختلفان, وهذا ما يساعدنا على اتخاذ قرار مناسب حول هذين المجتمعين .

ولإنشاء مجال ثقة للنسبة $\frac{\sigma_1^2}{\sigma_2^2}$ نفترض أننا سحبنا من المجتمعين عينتين عشوائيتين بحجمين n_1و n_2 فكان تبايناهما s_1^2 و s_2^2 على الترتيب (يراعى أن يكون $s_1^2 < s_2^2$ وإذا لم يكن ذلك نغير ترقيم المجتمعين, وذلك حتى تتوافق مع جداول F المختصرة , والتي تقتصر على قيم F الأكبر من الواحد) . والمطلوب أن نبحث عن حدي المجال $\left] d_1 \text{ و } d_2 \right[$ بحيث يكون لدينا:

$$P\left[d_1 < \frac{\sigma_1^2}{\sigma_2^2} < d_2 \right] = \beta = 1 - \alpha \qquad (103-2)$$

لنضرب أطراف المتراجحة بمقلوب النسبة $\frac{s_1^2}{s_2^2}$, فنجد أن الشرط السابق يأخذ الشكل التالي:

$$P\left[d_1 \frac{s_2^2}{s_1^2} < \frac{s_2^2 \sigma_1^2}{s_1^2 \sigma_2^2} < d_2 \frac{s_2^2}{s_1^2} \right] = \beta = 1 - \alpha \qquad (104-2)$$

وبكتابة المقدار الأوسط على الشكل :

$$\frac{s_2^2 \sigma_1^2}{s_1^2 \sigma_2^2} = \frac{\dfrac{(n_2-1)s_2^2}{\sigma_2^2(n_2-1)}}{\dfrac{(n_1-1)s_1^2}{\sigma_1^2(n_1-1)}} = F \quad \sim F_{v_2 v_1} \qquad (105-2)$$

فإننا نحصل على متحول F يخضع للتوزيع $F_{(n_2-1, n_1-1)}(x)$ (انظر النظريات في الاحتمالات) .

وبما إننا رمزنا إلى المتحول السابق بالرمز F فإن المتراجحة (2-104) تأخذ الشكل التالي :

$$P\left[d_1 \frac{s_2^2}{s_1^2} < \mathrm{F} < d_2 \frac{s_2^2}{s_1^2} \right] = 1 - \alpha$$

$$P\left[F < d_2 \frac{s_2^2}{s_1^2} \right] - P\left[F < d_1 \frac{s_2^2}{s_1^2} \right] = 1 - \alpha \qquad (106-2)$$

وبمقارنة هذه العلاقة مع تابع التوزيع التجميعي لـ F ,الذي يعطينا قيم F_P الحدية, المقابلة للاحتمال اليساري P, والذي يكون له الصيغة التالية :

$$P(F < F_P) = \int_0^{F_P} c x^{\frac{k_1}{2}-1} * (k_1 x + k_2)^{-\frac{k_1+k_2}{2}} * dx = P \qquad (107-2)$$

فإنه يمكننا أن نستنتج ونضع أن:

$$P\left(F < d_1 \frac{s_2^2}{s_1^2}\right) = P_1 \qquad , \qquad P\left(F < d_2 \frac{s_2^2}{s_1^2}\right) = P_2$$

وبذلك نحصل من(2−106) على المعادلة التالية:

$$P_2 - P_1 = 1 - \alpha \qquad (108-2)$$

وهي معادلة تتضمن مجهولين هما P_1 وP_2 , وبما أنها معادلة وحيدة بمجهولين فإن لها لا نهائية من الحلول . وحتى نحصل على حل ما لابد من أن نثبت أحد المجهولين ثم نحسب الآخر . ولذلك نثبت أحدهما, بحيث نترك احتمالين متساويين على طرفي المجال المطلوب ونضع ذلك كما يلي(انظر الشكل(2−5) اللاحق):

$$P\left[F < d_1 \frac{s_2^2}{s_1^2}\right] = \frac{\alpha}{2} = P_1 \qquad (109-2)$$

وبالتعويض في (2−108) نجد أن:

$$P\left[F < d_2 \frac{s_2^2}{s_1^2}\right] = 1 - \frac{\alpha}{2} = P_2 \qquad (110-2)$$

وبمقارنة هاتين العلاقتين مع العلاقة الجدولية (2−107) نستنتج أن:

$$d_1 \frac{s_2^2}{s_1^2} = F_{\frac{\alpha}{2}}(n_2 - 1 \,, n_1 - 1) \qquad (111-2)$$

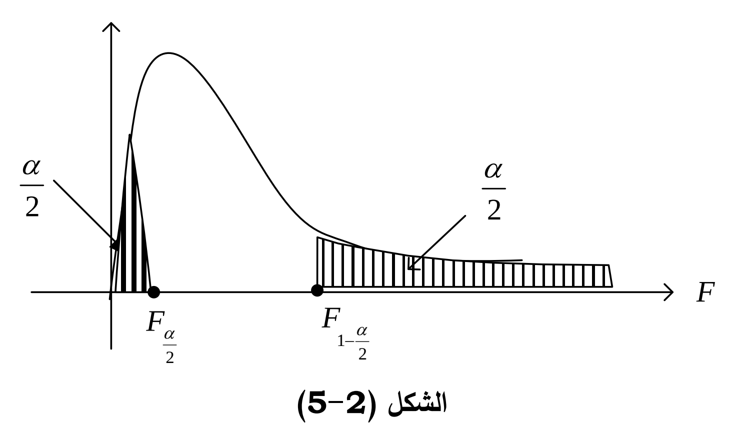
$$d_2 \frac{s_2^2}{s_1^2} = F_{1-\frac{\alpha}{2}}(n_2 - 1 \,, n_1 - 1) \qquad (112-2)$$

وذلك لأن النسبة F تخضع للتوزيع $F(x)$ ذي درجتي حرية $\left[n_2 - 1 \text{ و } n_1 - 1\right]$ المعكوستين, ومن ثم نجد أن الحدين المطلوبين يساويان:

$$d_1 = \frac{s_1^2}{s_2^2} * F_{\frac{\alpha}{2}}(n_2 - 1 \,, n_1 - 1) \qquad (113-2)$$

$$d_2 = \frac{s_1^2}{s_2^2} * F_{1-\frac{\alpha}{2}}(n_2 - 1 \,, n_1 - 1) \qquad (114-2)$$

وإن الشكل رقم (2−5) يوضح لنا معنى هاتين القيمتين لـ F ويبين موقعهما . ولإيجاد هاتين القيمتين من الجداول الإحصائية التراكمية, نقوم أولاً بإيجاد القيمة الحدية $F_{1-\frac{\alpha}{2}}$ من جداول F المناسبة للاحتمال أو لمستوى الدلالة المطلوب.



الشكل (2−5)

ونأخذ قيمتها من الحجرة المقابلة لدرجتي حرية $\left(n_1 - 1\right.$ و $\left.n_2 - 1\right)$ على الترتيب، مع ملاحظة أنهما معكوستين $\left(k_2 = n_1 - 1\right.$ و $\left.k_1 = n_2 - 1\right)$. علماً بأن الجداول الملحقة بهذا الكتاب تتضمن قيم F_P المقابلة للاحتمالات التراكمية : 0.90 أو 0.95 أو0.99 , أي المقابلة لمستويات الدلالة:

$$\frac{\alpha}{2} = 0{,}10, \qquad \frac{\alpha}{2} = 0{,}05, \qquad \frac{\alpha}{2} = 0{,}01,$$

ولكن حساب القيمة $F_{\frac{\alpha}{2}}$ الصغيرة (أصغر من الواحد) يحتاج إلى جداول تفصيلية, لذلك نستعين بخواص التوزيع F لحسابها من نفس جداول $\boldsymbol{F_{1-\frac{\alpha}{2}}}$, حيث نعلم أن:

$$F_{\frac{\alpha}{2}}(n_2 - 1\,, n_1 - 1\,) = \frac{\mathbf{1}}{\boldsymbol{F_{1-\frac{\alpha}{2}}(n_1-1\,,n_2-1)}} \qquad (115-2)$$

وبالتعويض في (2−113) نجد أن حدي المجال المطلوب يعطيان بالعلاقتين :

$$d_1 = \frac{s_1^2}{s_2^2} * \frac{1}{F_{1-\frac{\alpha}{2}}(n_1 - 1\,, n_2 - 1)} \qquad (116-2)$$

$$d_2 = \frac{s_1^2}{s_2^2} * F_{1-\frac{\alpha}{2}}(n_2 - 1\,, n_1 - 1) \qquad (117-2)$$

وبالتعويض في (2−103) نحصل على المجال التالي:

$$P\left[\frac{s_1^2}{s_2^2} * \frac{1}{F_{1-\frac{\alpha}{2}}(n_1 - 1\,, n_2 - 1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} * F_{1-\frac{\alpha}{2}}(n_2 - 1\,, n_1 - 1)\right] \qquad (118-2)$$

وهو المجال الذي يمكن أن يحوي النسبة $\frac{\sigma_1^2}{\sigma_2^2}$ باحتمال قدره $(1-\alpha)$ ويترك على كل من جانبيه احتمالاً قدره $\frac{\alpha}{2}$.

وأخيراً نشير إلى أن جميع قيم F الجدولية هي أكبر من الواحد لأنها مصممة على افتراض أن $s_1^2 > s_2^2$, وإذا لم يكن ذلك محققا فيجب تبديل رقمي المجتمعين حتى يتحقق ذلك الافتراض, وذلك حتى يتم استخدام الجداول بطريقة صحيحة .

- **مثال(2−9):** للتفضيل بين جهازين لقياس شدة التيار الكهربائي أجرينا بوساطة الأول *10* قياسات وبوساطة الثاني 8 قياسات، فوجدنا أن تبايني هاتين العينتين من القياسات يساويان $s_1^2 = 4{,}14$ و$s_2^2 =$ 3,21 , فأوجد مجال الثقة لنسبة تبايني الجهازين $\frac{\sigma_1^2}{\sigma_2^2}$ باحتمال ثقة قدره 0.90 .

الحل: نقوم بحساب حدى المجال $[d_1\,, d_2\,]$ من العلاقتين (2−116) و(2−117), فنلاحظ أن عددا درجتي الحرية يساويان: $n_2 - 1\ = 7 \qquad n_1 - 1 = 9$

وبما أن $\frac{\alpha}{2} = 0.05$, فإننا نجد من جدول F أن :

$$F_{1-\frac{\alpha}{2}}(n_2 - 1\,, n_1 - 1) = F_{0.95}(7\,, 9) = 3.29$$

$$F_{1-\frac{\alpha}{2}}(n_1 - 1\,, n_2 - 1) = F_{0.95}(9\,, 7) = 3.68$$

وبذلك نجد أن:

$$d_1=\frac{s_1^2}{s_2^2}*\frac{1}{F_{1-\frac{\alpha}{2}}(9\,,7)}=\frac{4.14}{3.21}*\frac{1}{3.68}=0.35$$

$$d_2=\frac{s_1^2}{s_2^2}*F_{1-\frac{\alpha}{2}}(7\,,9)=\frac{4.14}{3.21}*3.29=4.24$$

وبذلك نحصل على المجال التالي :

$$P\left(0.35<\frac{\sigma_1^2}{\sigma_2^2}<4.24\right)=0.90$$

وهو يميل إلى اليمين ولكنه يتضمن القيمة واحد التي تدل على تساوي التباينين .

2-9: الطريقة العامة لإيجاد مجالات الثقة (للعينات الكبيرة):

لنفترض أننا سحبنا عينة كبيرة بحجم n من مجتمع خاضع للتوزيع $f(x\,,\theta)$ الذي يتضمن مؤشراً وحيداً مجهولاً θ . ونريد أن ننشئ مجال ثقة باحتمال ثقة قدره $(1-\alpha)$ للمؤشر θ .

لقد رأينا في الفصل السابق أن توزيع تقدير الإمكانية العظمى $\tilde{\theta}$ لـ θ يتقارب من التوزيع الطبيعي($N(\theta\,,\sigma_{\tilde{\theta}}^2)$؛ حيث أن $\sigma_{\tilde{\theta}}^2$ اعتماداً على (48a-1) يساوي :

$$\sigma_{\tilde{\theta}}^2=\frac{1}{n\,E\left(-\frac{\partial^2 1n\,f}{\partial\theta^2}\right)}=\frac{1}{E\left(-\frac{\partial^2 1n\,L}{\partial\theta^2}\right)}\qquad(119-2)$$

وبناء على ذلك يمكننا أن نكتب مجال الثقة لـ θ على الشكل التالي:

$$P\left(-z_{1-\frac{\alpha}{2}}<\frac{\tilde{\theta}-\theta}{\sigma_{\tilde{\theta}}}<z_{1-\frac{\alpha}{2}}\right)=1-\alpha\qquad(120-2)$$

وبتعويض $\sigma_{\tilde{\theta}}$ بقيمته من العلاقة السابقة نحصل على المجال التالي:

$$P\left[\tilde{\theta}-\frac{z_{1-\frac{\alpha}{2}}}{\sqrt{E\left(-\frac{\partial^2 1n\,L}{\partial\theta^2}\right)}}<\theta<\tilde{\theta}+\frac{z_{1-\frac{\alpha}{2}}}{\sqrt{E\left(-\frac{\partial^2 1n\,L}{\partial\theta^2}\right)}}\right]=1-\alpha\qquad(121-2)$$

- **مثال(2-10)**: أوجد مجال الثقة باحتمال 0.95 للمؤشر λ في توزيع بواسون اعتماداً على عينة كبيرة بحجم n .

نعلم أن توزيع بواسون هو:

$$f(x\,,\lambda)=\frac{\lambda^x\,e^{-\lambda}}{x!}\qquad(x:0\,,1\,,2\,,\ldots..)$$

وأن تابع الامكانية العظمى له هو

$$L(x\,,\lambda)=\frac{\lambda^{\sum x_i}\,e^{-n\,\lambda}}{\prod_{i=1}^n x_i}$$

وإن معادلة الإمكانية العظمى هي:

$$\frac{\partial 1n\,L}{\partial\lambda}=\frac{n}{\lambda}(\bar{x}-\lambda)=0$$

وهي تعطينا أن: $\tilde{\lambda} = \bar{x}$

كما نجد أن :

$$\frac{\partial^2 1n\, L}{\partial \lambda^2} = \frac{-n\bar{x}}{\lambda^2}$$

$$E\left(\frac{-\partial^2 1n\, L}{\partial \lambda^2}\right) = E\left(\frac{+n\bar{x}}{\lambda^2}\right) = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$

وهكذا نجد أن مجال الثقة لـ λ حسب (2−121) هو :

$$P\left[\bar{x} - \frac{1.96}{\sqrt{\frac{n}{\lambda}}} < \lambda < \bar{x} + \frac{1.96}{\sqrt{\frac{n}{\lambda}}}\right] = 0.95$$

ولاستخلاص λ من حدي مجال الثقة نكتب حدي المجال ونعالجهما كما يلي:

$$\lambda = \bar{x} \pm \frac{1.96\sqrt{\lambda}}{\sqrt{n}}$$

$$(\lambda - \bar{x})^2 = 3.84 * \frac{\lambda}{n}$$

$$\lambda^2 - 2\lambda\bar{x} + \bar{x}^2 = 3.84\frac{\lambda}{n}$$

$$\lambda^2 - \left(2\bar{x} + \frac{3.84}{n}\right)\lambda + \bar{x}^2 = 0$$

وبحلها نجد أن:

$$\lambda = \bar{x} + \frac{1.92}{n} \pm \sqrt{\bar{x}\frac{3.84 * 2}{n} + \frac{3.69}{n^2}}$$

أي أن مجال الثقة المطلوب هو:

$$P\left[\bar{x} + \frac{1.92}{n} - \sqrt{\bar{x}\frac{3.84 * 2}{n} + \frac{3.69}{n^2}} < \lambda < \bar{x} + \frac{1.92}{n} + \sqrt{\bar{x}\frac{3.84 * 2}{n} + \frac{3.69}{n^2}}\right] = 0.95$$

وهنا نلاحظ أن هذا المجال غير متناظر حول $\bar{x}$ بل حول $\bar{x} + \frac{1.92}{n}$.

ملاحظة هامة:

من المهم جداً أن نشير إلى أن جميع التقديرات السابقة ومجالات الثقة المتعلقة بها، محسوبة بافتراض أن الظواهر أو الخواص خاضعة للتوزيع الطبيعي العام . وأن مسائل التقدير تصبح أكثر تعقيداً عندما تكون الخاصة غير خاضعة للتوزيع الطبيعي . إلا أن الخروج المعتدل عن هذا الشرط لا يؤثر بشكل جدي في أمثال التقدير $\left(z_{1-\frac{\alpha}{2}}\right.$ أو $\left.t_{\frac{\alpha}{2}}\right)$.

كما أنه إذا كان حجم العينة كبيراً $(n \geq 30)$ فإنه يمكننا الاعتماد على التوزيع الطبيعي لإيجاد مجال الثقة لكل من μ و $(\mu_1 - \mu_2)$ و R و $R_1 - R_2$ ...الخ مهما كان توزيعها الأساسي . والخطأ المرتكب في هذه الحالات يمكن إهماله .

تمارين الفصل الثاني

1−لدراسة وزن قطع الزبدة المنتجة في معملين A وB سحبنا عينة من كل منهما بحجمين $n_1 = 10$ و $n_2 = 12$ فوجدنا أن متوسطي وزن القطعة يساويان : $\bar{x}_1 = 240\ g$ ، $\bar{x}_2 = 250\ g$
أما تبايناهما فكانا يساويان : $s_1^2 = 5000$ ، $s_2^2 = 4000$ أوجد مجال ثقة باحتمال 0.95 لكل من المؤشرات التالية :

- متوسط المجتمع الأول μ_1 .
- متوسط المجتمع الثاني μ_2 .
- الفرق بين المتوسطين $(\mu_1 - \mu_2)$ وبفرض أن $\sigma_1^2 = \sigma_2^2$.
- تباين المجتمع الأول σ_1^2 .
- تباين المجتمع الثاني σ_2^2 .
- نسبة التباينين σ_1^2/σ_2^2 .

2− عند رمي قطعة نقود 400 مرة حصلنا على الصورة 175 مرة . أوجد مجال الثقة ذا الاحتمال 0.90 لاحتمال ظهور الصورة ؟ ثم أوجد مجال الثقة ذي الاحتمال 0.99 لاحتمال ظهور الصورة ؟
لنفترض أننا أجرينا 500 تجربة أخرى على القطعة النقدية نفسها فحصلنا على الصورة 230 مرة .
فأوجد مجال الثقة ذا الاحتمال 0.90 لاحتمال ظهور الصورة ؟
ثم ادرس هل الفرق بين الاحتمالين جوهري وهل القطعة النقدية متوازنة ؟

3− لقياس السرعة القصوى لطائرة جديدة أجرينا 15 تجربة فحصلنا على متوسط للسرعة القصوى يساوي $\bar{x} = 424{,}7\ M/C$ وانحراف معياري لها $S = 8{,}7\ M/C$.
والمطلوب : − إيجاد مجال الثقة لكل من متوسط السرعة وتباينها باحتمال قدره $\beta = 0.90$.
- حساب احتمال أن تكون القيمة المطلقة للخطأ في تقدير المتوسط أقل من $2\ M/C$.
(بفرض أن القياسات خاضعة للتوزيع الطبيعي) .

4− إذا كان متوسط الخطأ في جهاز لقياس ارتفاع الطائرة هو 15 م, فكم جهازاً يجب أن نضع على الطائرة حتى نجعل الخطأ في متوسط الارتفاع $\bar{x} - \mu$ أصغر من (30+) م وذلك باحتمال قدره 0.99 وبفرض أن الخطأ يخضع للتوزيع الطبيعي .

5− في اختبار على 16 مصباحاً كهربائياً وجدنا أن تقدير متوسط عمر المصباح كان يساوي $\bar{x} = 3000\ H$ وأن تباينه $s^2 = 4000\ H^2$ وبفرض أن عمر المصباح يخضع للتوزيع الطبيعي فأوجد:
- مجال الثقة لكل من متوسط عمر المصباح μ وتباينه σ^2 باحتمال قدره 0.90 ؟

6−أوجد مجال الثقة باحتمال قدره 0.90 للوسيط P في التوزيع الثنائي وذلك اعتماداً على الطريقة العامة لإيجاد مجالات الثقة .

الفصل الثالث : اختبارات الفرضيات البسيطة

3-1: تمهيد :

تعتمد اختبارات الفرضيات على البيانات المتوفرة عن الظواهر المدروسة في المجتمعات الإحصائية. والبيانات هي قيم عددية أو حالات وصفية تعبر عن المتحولات المعرفة على الظاهرة المدروسة . وبذلك ويمكننا تصنيف البيانات إلى نوعين أساسيين هما:

أ- بيانات كمية: وهي بيانات عددية عن متحولات قابلة للقياس بواحدات قياس محددة، وهذه البيانات يمكن أن تكون:

- منقطعة: كعدد أفراد الأسرة- وعدد الطلاب- وعدد السيارات ...الخ .
- مستمرة: كعمر الإنسان- درجة الحرارة- مقدار الدخل ....الخ .

ب- بيانات نوعية: وهي حالات وصفية لمتحولات غير قابلة للقياس، وهذه المعلومات يمكن أن تكون:

- أسمية: كحالات الجنس- حالات العمل- الحالة الاجتماعية ...الخ .
- مرتبة: كحالات التعليم- حالات الوظيفة- حالات الرضا ..الخ .

ويتم تجميع البيانات عن الظاهرة المدروسة أو عن المتحولات المطلوبة من عناصر المجتمع الاحصائي بواسطة أحد الأسلوبين :

- الحصر الشامل: وهو يشمل جميع عناصر المجتمع الاحصائي المؤلف من N عنصراً
- المسح بالعينة: وهو يشمل جزء من المجتمع ويكون على شكل عينة حجمها n عنصراً, تسحب عشوائياً من عناصر ذلك المجتمع بدون إعادة أو مع الاعادة .

وتستخدم بيانات هذه العينة لتقدير معالم المجتمع المجهولة مثل: المتوسط μ أو التباين σ^2 أو النسبة فيه R، وذلك من خلال استخدام المؤشرات المقابلة لها والمحسوبة من العينة، والتي سنسميها (مؤشرات العينة)، وهي متوسط العينة $\bar{x}$ وتباين العينة S^2 والنسبة في العينة r . ويبُرهن في نظرية العينات أن مؤشرات العينة المصححة هي تقديرات غير متحيزة وفعالة ومتماسكة لمعالم المجتمع المقابلة لها .

والآن لنفترض أننا نقوم بدراسة خصائص أحد المتحولات الكمية X من عناصر المجتمع المدروس (وليكن X وزن الطالب في الجامعة)، لذلك نسحب عينة عشوائية من طلاب ذلك المجتمع بحجم n طالباً , فنحصل من كل طالب i فيها على وزن محدد x_i , ويكون لدينا القياسات التالية :

$$X: x_1 \quad x_2 \quad x_3 \dots\dots . x_i \dots\dots x_n$$

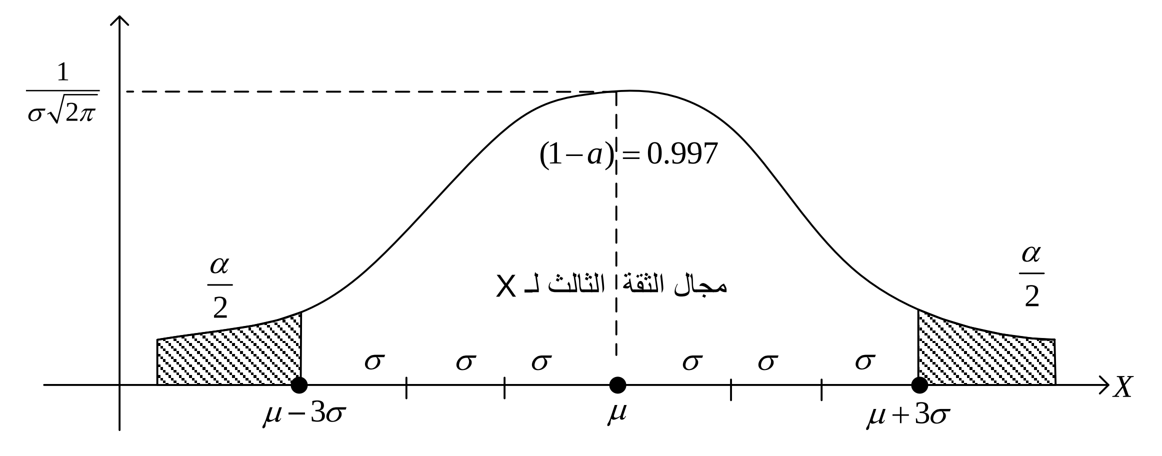
وبناء على نظرية العينات نقوم بتقدير معالم المجتمع المجهولة من مؤشرات العينة المعلومة وحساب مقدار الخطأ المرتكب في كل تقدير وفق الجدول التالي :

جدول (3-1): معالم المجتمع وتقديراتها من مؤشرات العينة :

معالم المجتمع المجهولة للمتحول	تقديرات المعالم من مؤشرات العينة
1-قيمة المتوسط في المجتمع μ	تقدر من متوسط العينة: $\bar{x} = \frac{1}{n}\sum x_i$ ونكتب ذلك كما يلي: $\tilde{\mu} = \bar{x}$
2-قيمة تباين المجتمع σ^2	تقدرمن تباين العينة المصحح والمعرف بالعلاقة: $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ ونكتب ذلك كما يلي: $\tilde{\sigma}^2 = s^2$
3-قيمة النسبة في المجتمع R للمتحول الثنائي (0 او 1)	تقدر من النسبة في العينة $r = \frac{m}{n}$ ونكتب ذلك كما يلي: $\tilde{R} = r$ حيث m عدد الظهور و n حجم العينة وتباينها يساوي $s^2 \approx r(1-r)$
4-الخطأ المعياري المرتكب في تقدير متوسط المجتمع في حالة السحب مع الاعادة $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$	يقدر من خلال ما يقابله في العينة : $\tilde{\sigma}_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$
5-الخطأ المعياري المرتكب في تقدير المتوسط في حالة السحب بدون إعادة : $\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1} * \frac{\sigma^2}{n}}$	يقدر من خلال العينة بالعلاقة: $\tilde{\sigma}_{\bar{x}} = \sqrt{\frac{N-n}{N-1} * \frac{s^2}{n}} \approx \sqrt{\frac{N-n}{N} * \frac{s^2}{n}}$
6-الخطأ المعياري المرتكب في تقدير النسبة R في حالة السحب مع الاعادة: $\sigma_r = \sqrt{\frac{R(1-R)}{n}}$	يقدر من خلال خطأ النسبة r في العينة كمايلي: $\tilde{\sigma}_r = \sqrt{\frac{r(1-r)}{n}}$
7-الخطأ المعياري المرتكب في تقدير النسبة R في حالة السحب بدون إعادة : $\sigma_r = \sqrt{\frac{N-n}{N-1} * \frac{R(1-R)}{n}}$	يقدر من خلال خطأ النسبة r في العينة كما يلي: $\tilde{\sigma}_r = \sqrt{\frac{N-n}{N-1} * \frac{r(1-r)}{n}}$ $= \sqrt{\frac{N-n}{N} * \frac{r(1-r)}{n}}$

ولكن عملية التقدير لا تنتهي عند ذلك , بل يجب إنشاء مجال ثقة يحتوي المعلم الذي نقدره في المجتمع. وحتى نستطيع إنشاء مجال ثقة يجب أن يكون التوزيع الاحتمالي للمتحول X معلوماً. ومنه يجب أن يكون توزيع متوسط العينة $\bar{x}$ معلوماً أيضاً.

واختصاراً لهذه القضايا نفترض أن المتحول المدروس X يخضع في المجتمع للتوزيع الطبيعي العام $N(\mu , \sigma^2)$ الذي متوسطه μ وتباينه σ^2 وهو يرسم الشكل التالي :



الشكل (3−1) منحنى التوزيع الطبيعي للمتحول X ومجال الثقة الثالث

وبناءً على خواص التوزيع الطبيعي يمكننا إنشاء مجال الثقة لـ X المقابل لاحتمال الثقة $(1-\propto)$ أو لمستوى دلالة $(\propto)$ بحيث يكون:

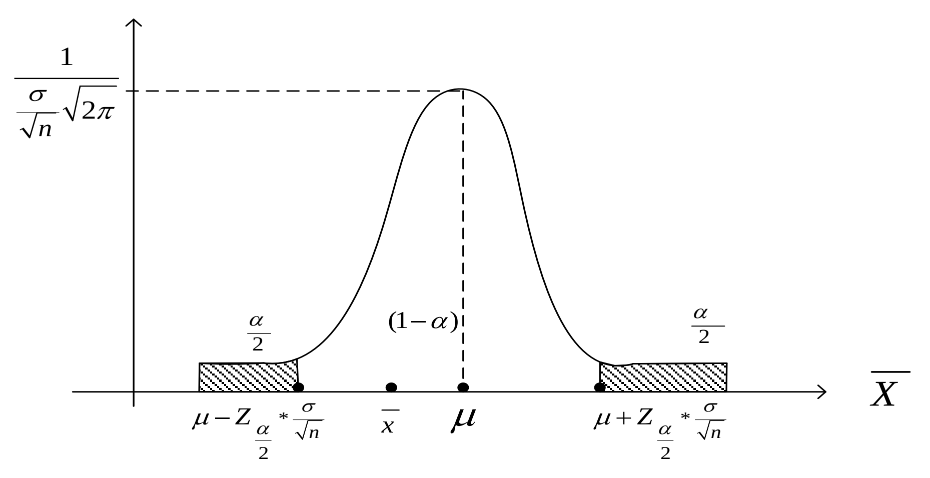
$$P\left[\mu - Z_{\frac{\propto^*}{2}}\sigma \leq X \leq \mu + Z_{\frac{\propto^*}{2}}\sigma\right] = 1-\propto \qquad (1-3)$$

حيث أن $Z_{\frac{\propto}{2}}$ هي قيمة متحول التوزيع الطبيعي المعياري التي تترك نصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$ على يمينها و$\left(\frac{\propto}{2}\right)$ على يسار $-Z_{\frac{\propto}{2}}$, [حيث استبدلنا الرمز $Z_{1-\frac{\propto}{2}}$ بالرمز $Z_{\frac{\propto}{2}}$ للاختصار]. ولقد أنشأنا على الشكل (1−1) مجال الثقة الثالث الذي يقابل احتمال ثقة (0,997).

ملاحظة هامة: لقد استبدلنا الرمز $Z_{1-\frac{\propto}{2}}$ بالرمز $Z_{\frac{\propto}{2}}$ للاختصار, وهو يساوي قيمة Z التي تترك على يمينها $\frac{\propto}{2}$.

وكذلك فإننا سنقوم باستبدال الرموز $t_{1-\frac{\alpha}{2}}$ و $\chi_{1-\frac{\alpha}{2}}$ و $F_{1-\frac{\alpha}{2}}$ بالرموز $t_{\frac{\alpha}{2}}$ و $\chi_{\frac{\alpha}{2}}$ و $F_{\frac{\alpha}{2}}$ لنفس الهدف ولنفس المعنى, وهذا مايستدعي الإنتباه عند استخراج القيم الحرجة لهذه المتحولات من الجداول الإحصائية.

وعندما نسحب عينة من عناصر المجتمع بحجم n نحصل منها على متوسط هو $\bar{x}$ وعلى تباين هوS^2، ولكن هذه العينة ليست وحيدة بل يمكن أن يسحب غيرنا وغيرنا عينات أخرى، فيحصل على متوسطات أخرى $\bar{x}_k$ وتباينات أخرى S_k^2، وبما أن عدد العينات الممكنة يساوي C_N^n عينة (في حالة السحب بدون إعادة)، فإن هذا يعني أنه يمكننا أن نحصل على C_N^n متوسطاً $\bar{x}_k$، وكل منها يعتبر تقديراً لمتوسط المجتمع μ . وبناءً عليه يكون متوسط العينة $\bar{x}$ هو الآخر متحولاً عشوائياً جديداً متوسطه μ أيضاً، ولكن تباينه يساوي $\frac{\sigma^2}{n}$ (وانحرافه المعياري يساوي $\frac{\sigma}{\sqrt{n}}$) ويخضع للتوزيع الطبيعي العام $N\left(\mu , \frac{\sigma^2}{n}\right)$، والذي يأخذ الشكل الضامر والمتطاول التالي:



الشكل (3−2): توزيع متوسط العينة $\bar{x}$

وبناءً على شكل التوزيع (3−2) وقياساً على العلاقة (3−1) يمكننا أيضاً إنشاء مجال الثقة للمتوسط $\bar{x}$ المقابل لاحتمال الثقة (∝−1) أو لمستوى الدلالة (∝) بحيث يكون:

$$P\left[\mu - Z_{\frac{\propto}{2}} * \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + Z_{\frac{\propto}{2}} * \frac{\sigma}{\sqrt{n}}\right] = 1-\propto \qquad (2-3)$$

وحيث أن $Z_{\frac{\propto}{2}}$: هي القيمة الجدولية (الحرجة) لمتحول التوزيع الطبيعي المعياري المقابلة لنصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$ على اليمين و $\left(\frac{\propto}{2}\right)$ على اليسار. وهذا المجال يضمن لنا أن تقع قيمة $\bar{x}$ ضمنه باحتمال يساوي $(1-\propto)$.

وحتى نستفيد من العلاقة (1-2) نطرح من أطرافها μ ثم نقسمها على $\frac{\sigma}{\sqrt{n}}$ فنحصل على ما يلي:

$$P\left[-Z_{\frac{\propto}{2}} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq Z_{\frac{\propto}{2}}\right] = 1-\propto \qquad (3-3)$$

وهذا المجال يضمن لنا أن تقع قيمة المقدار $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$ ضمن المجال $\left[-Z_{\frac{\propto}{2}}\ ,+Z_{\frac{\propto}{2}}\right]$ باحتمال يساوي $(1-\propto)$، وهنا نلاحظ أن المقدار $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$ هو متحول عشوائي ثالث، وهنا نميز بين حالتين هما:

أ- إذا كانت قيمة σ^2 وبالتالي قيمة σ معلومة من المجتمع: فإن المقدار $\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$ يخضع للتوزيع الطبيعي المعياري $N(0\ ,1)$ لأنه ناتج عن معايرة المتحول المتوسط $\bar{x}$، لذلك نرمز له بـ Z ونكتبه كما يلي:

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \qquad (4-3)$$

وهكذا نجد أنه يمكننا اعتبار هذا المقدار مؤشراً لاختبار أي فرضية حول متوسط المجتمع μ (مثل الفرضية $H_0: \mu = \mu_0$)، وذلك عندما نعوض μ بـ μ_0 ونقوم بحساب قيمة Z من العلاقة (3−4) وبشرط أن تكون قيمة σ معلومة .

ثم نقارن قيمة Z المحسوبة مع طرفي المجال المعرف في [3−1] وهو $\left[-Z_{\frac{\propto}{2}}\ ,+Z_{\frac{\propto}{2}}\right]$. ونتخذ القرار كمايلي : إذا كانت قيمة Z المحسوبة واقعة ضمن ذلك المجال نقبل تلك الفرضية $(H_0: \mu = \mu_0)$ ، أما إذا كانت قيمة Z المحسوبة واقعة خارجه من الطرفين، فإننا نرفض الفرضية $(H_0: \mu = \mu_0)$، وبعبارة أخرى

إذا كانت $|Z| \leq Z_{\frac{\propto}{2}}$ نقبل الفرضية المذكورة ، وإذا كانت $|Z| > Z_{\frac{\propto}{2}}$ نرفض الفرضية السابقة. علماً بأن احتمال قبول الفرضية $(H_0: \mu = \mu_0)$ يساوي $(1-\propto)$ واحتمال رفضها من الطرفين يساوي $(\propto)$.

ب-أما إذا كانت قيمة σ^2 في المجتمع مجهولة, فإننا نلجأ إلى تقديرها من خلال تباين العينة المصححS^2 , ونقوم باستبدال قيمة σ في (3-4) بتقديرها S من العينة، فنحصل على متحول عشوائي جديد مركب من متحولين عشوائيين $\bar{x}$ وS ونرمز له بـ t ونكتبه كما يلي:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad (5-3)$$

ومعلوم من نظرية الاحتمالات والإحصاء الرياضي أن المتحول t يخضع لتوزيع (ستودينت) ذي(n-1) درجة حرية. (وهو توزيع يتقارب مع التوزيع الطبيعي المعياري عندما تصبح $n \geq 30$) . وقياساً على العلاقة (3-3) يمكننا أن ننشأ مجال الثقة للمتحول t كما يلي:

$$P\left[-t_{\frac{\propto}{2}} \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq +t_{\frac{\propto}{2}}\right] = 1-\propto \qquad (6-3)$$

حيث أن $t_{\frac{\propto}{2}}$ هي قيمة متحول (ستودينت) الجدولية (أوالحرجة), المقابلة لنصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$ على اليمين ولـ $\left(\frac{\propto}{2}\right)$ على اليسار ولـ (n-1) درجة حرية، وهذا المجال يضمن لنا أن تقع قيمة t ضمن المجال $\left[-t_{\frac{\propto}{2}}\ , +t_{\frac{\propto}{2}}\right]$ باحتمال قدره $(1-\propto)$. وهكذا نجد أنه يمكننا الاستفادة من المتحول t في حالة العينات الصغيرة في اختبار أي فرضية حول متوسط المجتمع μ (مثل $H_0: \mu = \mu_0$) فنعوض μ بـ μ_0 ونحسب قيمة t من العلاقة (3-5) , ثم نقارنها مع طرفي المجال المعرف في (3-6) وهو المجال $\left[-t_{\frac{\propto}{2}}\ ,\ +t_{\frac{\propto}{2}}\right]$, ونتخذ القرار كما يلي:

إذا كانت قيمة t المحسوبة واقعة ضمن ذلك المجال تقبل تلك الفرضية $(H_0: \mu = \mu_0)$، أما إذا كانت t المحسوبة واقعة خارجه فإننا نرفض الفرضية $(H_0: \mu = \mu_0)$.

وبعبارة أخرى: إذا كانت $|t| \leq t_{\left(\frac{\propto}{2}, n-1\right)}$ فإننا نقبل الفرضية $(H_0: \mu = \mu_0)$، أما إذا كانت $|t| > t_{\left(\frac{\propto}{2}, n-1\right)}$ فإننا نرفض الفرضية المذكورة. علماً بأن احتمال قبول الفرضية $(\mu = \mu_0)$ يساوي $(1-\propto)$ واحتمال رفضها من الطرفين يساوي $(\propto)$.

وهكذا نجد أنه يمكننا، وباتباع نفس الأسلوب، استنباط العديد من المؤشرات لاستخدامها في اختبارات الفرضيات المختلفة (كل حالة حسب طبيعتها وحسب توزيعها الاحتمالي) , فنحصل على مؤشرات لتقدير النسبة R والتباين σ^2 وغيرهما.

ومن جهة أخرى يمكننا أن ننشأ مجال ثقة يحوي متوسط المجتمع μ وذلك بمعالجة العلاقة (3-2) وطرح المقدار $(\bar{x} - \mu)$ من أطرافها فنحصل على المجال التالي:

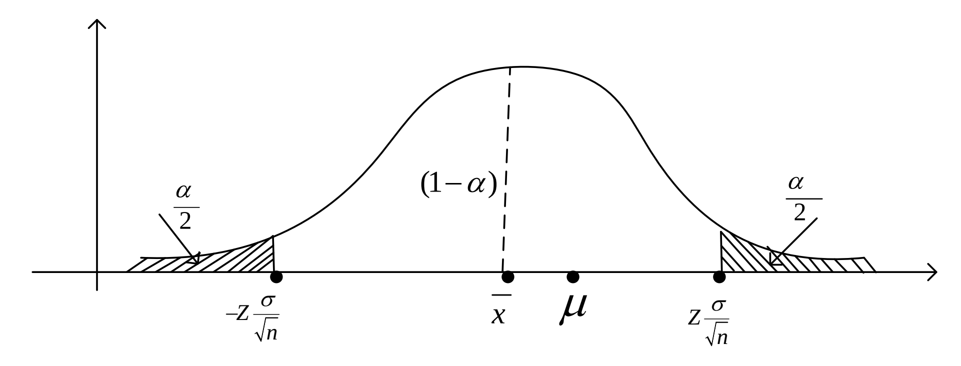
$$P\left[\bar{x} - Z_{\frac{\propto}{2}} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\frac{\propto}{2}} * \frac{\sigma}{\sqrt{n}}\right] = 1-\propto \qquad (7-3)$$

وهو مجال مركزه متوسط العينة $\bar{x}$ (وليس μ) ونصف طوله يساوي $\left(Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right)$, ويضمن لنا أنه يحتوي على متوسط المجتمع المجهول μ باحتمال $(1-\alpha)$ والشكل (3-3) يوضح ذلك .

وإذا كان التباين σ^2 مجهولاً فإننا نستبدله بتقديره S^2، وعندها فإن مجال الثقة (3-7) يصبح معرفاً على توزيع (ستودينت) ذي $(n-1)$ درجة حرية ويأخذ الشكل التالي:

$$P\left[\bar{x} - t_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}\right] = 1-\alpha \qquad (8-3)$$

وهو مجال مركزه $\bar{x}$ ونصف طوله $\left(t_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}\right)$, ويضمن لنا أنه يحتوي على متوسط المجتمع المجهول μ باحتمال $(1-\alpha)$.



الشكل (3-3) مجال الثقة لـ μ

3-2: أنواع وأسماء أهم الاختبارات :

توضع الفرضيات على معالم المجتمع (مثل μ أو R أو σ^2) أو على بعض خصائصه (مثل: الاستقلال، التوافق، التكرار، الالتواء ...الخ) وتختبر باستخدام معلومات العينة ومؤشرات الاختبار.

وتصنف الاختبارات إلى نوعين أساسيين :

أ- الاختبارات المعلمية: وتطبق على المتحولات الكمية .

ب- الاختبارات اللامعلمية: وتطبق على المتحولات النوعية والرتبية.

كما يمكن تصنيف الاختبارات حسب عدد المجتمعات والعينات كما يلي :

1- اختبارات لمجتمع واحد (عينة واحدة) .
2- اختبارات لمجتمعين (عينتين مستقلتين) .
3- اختبارات لعدة مجتمعات (لعدة عينات مستقلة) .
4- اختبارات لعينتين مترابطتين
5- اختبارات لعدة عينات مترابطة .

ويتضمن الجدول التالي أهم الاختبارات المعلمية واللامعلمية ومجالات تطبيقها

جدول (3-2): أهم الاختبارات الاحصائية المعلمية واللامعلمية:

أهم الاختبارات المعلمية	أهم الاختبارات اللامعلمية
1- اختبار Z الطبيعي لعينة واحدة وهو يطبق على متوسط المجتمع μ وعلى النسبة R فيه، مثل العلاقة (1-4)	**1- اختبار χ^2 : للاستقلال والارتباط بين متحولين نوعيين أو أحدهما نوعي . في عينة واحدة (غير مرتبة)**
2- اختبار (ستودينت) t لعينة واحدة صغيرة وهو يطبق على μ وعلى R مثل العلاقة (1-5)	**2- اختبار Gamma: للاستقلال والارتباط بين متحولين مرتبين من عينة واحدة**
3- اختبار χ^2 لعينة واحدة ويطبق على تباين المجتمع σ^2	**3- اختبارات الثبات والصدق ويستخدم في الاستبيانات (للمعلومات المرتبة)**
4- اختبار Z الطبيعي لعينتين مستقلتين ويطبق على الفرق بين متوسطي المجتمعين أو على الفرق بين النسبتين فيهما مثل العلاقة (1-24)	**4- اختبار مكنمارا: للحالات غير المرتبة (جدول رباعي)**
5- اختبار (ستودينت) t لعينيتين مستقلتين ويطبق على الفرق $\mu_1 - \mu_2$ وعلى الفرق $R_1 - R_2$، مثل العلاقة (1-25)	**5- اختبار: ويلكوكسن للمتحولات المرتبة**
6- اختبارF لعينتين مستقلتين ويطبق على تبايني مجتمعين $\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$، مثل العلاقة (1-35)	**6- اختبار كروسكال: للمتحولات المرتبة**
7- اختبار t للفرق بين الأزواج المتقابلة (عينتين مترابطتين) مثل العلاقة (1-54)	**7- اختبار مان ويتني: للمتحولات المرتبة**
8- اختبار تحليل التباين الأحادي لأكثر من عينتين مستقلتين، مثل العلاقة (1-40)	**8- اختبار t_B كيندال: للمتحولات المرتبة**
9- اختبار تحليل التباين الثنائي لمؤشرين على عدة مجتمعات .	**9- اختبار t_C كيندال: للمتحولات المرتبة**
10-اختبار χ^2 لتوافق التوزيعات الاحتمالية من عينة واحدة	**10-اختبار الاشارة للمتحولات الثنائية (1 ,0)**
11-اختبار كولموغوروف- سميرنوف لتوافق التوزيعات الاحتمالية	**11-اختبار كوكران- مينتال للبيانات المرتبة**
12-اختبار معنوية معامل الارتباط الخطي (معامل بيرسون)	**12-اختبار معنوية معامل الارتباط الرتبي (سبيرمان) للمتحولات المرتبة**

3-2-1: حساب موثوقية وقوة الاختبار

في الحقيقة إن عملية إجراء اختبار لأية فرضية عدم H_0، تتأثر بعدة عوامل أهمها حقيقة الفرضية H_0 في المجتمع ونوع القرار المتخذ بشأنها ، وبذلك نجد أنه لدينا الحالات التالية:

- إن فرضية العدم H_0 قد تكون بحقيقتها صحيحة أو خاطئة .
- إن القرار الذي سنأخذه حول H_0 يمكن أن يكون قبولاً أو رفضاً لها .

ويمكن وضع تقاطعات هذه الحالات الأربع في جدول كالتالي:

جدول(3-2a) : حالات تقاطع حقيقة فرضية العدم مع نوع القرار المتخذ حولها

نوع القرار المتخذ / حقيقة الفرضية H_0	قبول	رفض
H_0 صحيحة	القرار صحيح واحتماله =∝−1	القرار غير صحيح واحتماله =∝
H_0 خاطئة	القرار غير صحيح واحتماله = β	القرار صحيح واحتماله = $1-\beta$

ومن الجدول السابق نلاحظ إنه عندما نتخذ القرار حول H_0 , فيمكن أن يكون قرارنا غير صحيح في الحالتين التاليتين: رفض الفرضية الصحيحة، قبول الفرضية الخاطئة، وعندها سنرتكب أحد الخطأين التاليين:

- خطأ النوع الأول error type I: وهو قرار رفض الفرضية H_0 رغم إنها صحيحة، وإن احتمال وقوعنا في هذا الخطأ يسمى مستوى الدلالة ∝، ويسمى الاحتمال المتمم له (∝−1) بدرجة الثقة أو بالموثوقية .
- خطأ النوع الثاني error type II : وهو قرار قبول الفرضية H_0 رغم إنها خاطئة أو غير صحيحة، وإن احتمال وقوعنا في هذا الخطأ يساوي عدداً آخراً β، ويسمى الاحتمال المتمم له $(1-\beta)$ بقوة الاختبار .

وبناءً على ذلك تم تعريف قوة الاختبار: بأنها احتمال رفض الفرضية H_0 عندما تكون خاطئة. وهو يتمم احتمال قبولها β، أي أن قوة الاختبار تعرف بالاحتمال المتمم لـ β وهو يساوي :

$$W = 1 - \beta \qquad (3-8a)$$

ويتم حساب قيم β من تكاملات شرطية معقدة لا مجال للخوض فيها في هذا الفصل.

3-3: اختبارات معالم مجتمع طبيعي (من عينة واحدة):

3-3-1: اختبار متوسط المجتمع μ (أو النسبة R):

ويتألف من الخطوات التالية:

1- نحدد مستوى الدلالة (∝) أو احتمال الثقة (∝−1), وعادة يتم وضعه (∝= 0.05) أو (∝= 0.01 أو (∝= 0.10) .

2- نضع على متوسط المجتمع μ (أو النسبة R فيه) فرضيتين متنافيتين ومتكاملتين كما يلي:

أ- فرضية العدم: نفترض أن متوسط المجتمع μ يساوي قيمة معلومة μ_0، وهذا يعني أنه لا يوجد فرق معنوي بينه وبين القيمة المفترضة μ_0 (أي عدم وجود فرق بينهما) ونكتب ذلك كما يلي:

$$H_0: \mu = \mu_0 \qquad (9-3)$$

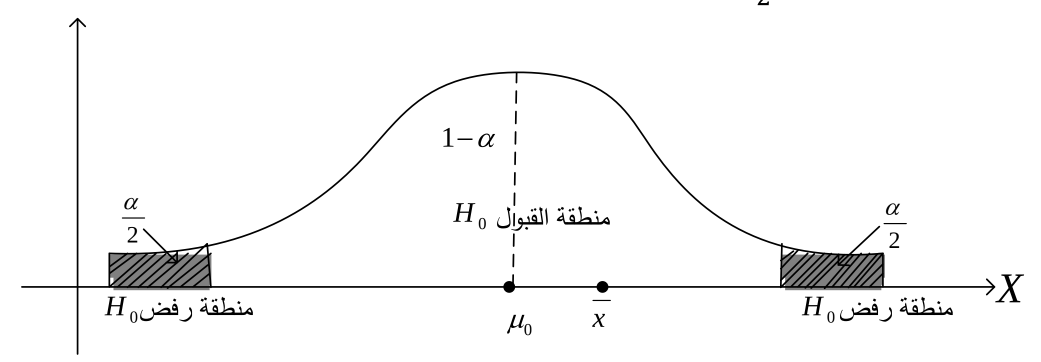
وتكون هذه الفرضية مقبولة إذا كان الفرق $(\mu - \mu_0)$ أو تقديره $(\bar{x} - \mu_0)$ واقعاً ضمن مجال الثقة المحدد للفرق $(\bar{x} - \mu)$, ونقرر ذلك من خلال مؤشر الاختبار المناسب.

ب- الفرضية البديلة: وهي الفرضية المعاكسة لفرضية العدم، ومن شكلها تتحدد منطقة الرفض, ويمكن أن تُكتب على أحد الأشكال الثلاثة التالية:

- الشكل الأول: الشكل الثنائي أو شكل عدم التساوي وتكتب الفرضية البديلة فيه كما يلي:

$$H_1: \mu \neq \mu_0 \qquad (10-3)$$

وفيه تكون منطقة الرفض واقعة على الجانبين، ولذلك يسمى هذا الشكل بالاختبار ثنائي الجانب، لأنه يخصص لكل جانب نصف مستوى الدلالة($\frac{\propto}{2}$) , كما في الشكل التالي :



الشكل (3-4) منطقة القبول ومنطقتي الرفض على اليمين واليسار

- الشكل الثاني: الأحادي اليميني، وتكتب الفرضية البديلة H_1 كما يلي:

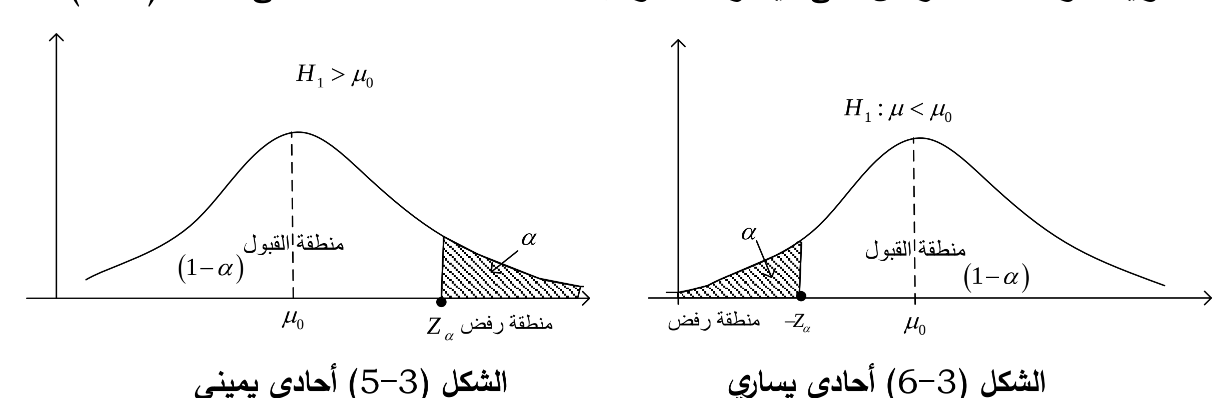
$$H_1: \mu > \mu_0 \qquad (11-3)$$

وفيه تكون منطقة الرفض على اليمين فقط، وتقابل كامل الاحتمال $\propto$ كما على الشكل (3-5).

- الشكل الثالث: الأحادي اليساري، وتكتب الفرضية البديلة H_1 كما يلي:

$$H_1: \mu < \mu_0 \qquad (12-3)$$

- وفيه تكون منطقة الرفض على اليسار فقط، وتقابل كامل الاحتمال $\propto$ كما على الشكل (3-6).



الشكل (3-5) أحادي يميني **الشكل (3-6) أحادي يساري**

وللتحقق من صحة أو عدم صحة فرضية العدم H_0، يجب علينا أن نسحب عينة عشوائية من المجتمع ونحسب متوسطها $\bar{x}$ ثم نقارنه مع متوسط المجتمع المفترض في الفرضية H_0 وهو μ_0. فإذا كان $\bar{x}$ يساويه أو قريباً منه نقبل فرضية العدم H_0، ونعترف بأن متوسط المجتمع يساوي μ_0، أما إذا كان $\bar{x}$ بعيداً عن μ_0 (يوجد فرق جوهري بينهما)، فإننا نرفض فرضية العدم H_0 ونقبل الفرضية H_1، ونعترف بأن متوسط المجتمع μ لايساوي μ_0، بل يساوي قيمة أخرى أكبر أو أصغر منها. وحتى لا تكون الأمور مزاجية فإن عملية مقارنة $\bar{x}$ مع μ_0، تحتاج إلى أداة إحصائية ورياضية تحدد لنا مقدار الفرق المقبول ومقدار الفرق المعنوي أو الجوهري، وهذه الأداة تسمى مؤشر الاختبار. وهكذا نجد أنفسنا بحاجة قبل كل شيء إلى سحب عينة عشوائية من المجتمع المدروس وحساب مؤشراتها المختلفة .

3−نقوم بتحديد طريقة سحب العينة (مع الإعادة أم بدون إعادة) , وحسب طريقة السحب المختارة نقوم بحساب حجم العينة من إحدى العلاقتين الخاصتين بتقدير المتوسط (أو النسبة R ضمن القوسين) وهما:.

$$n = \frac{Z^2 s^2}{d^2} = \left(\frac{Z^2 * r(1-r)}{d^2}\right) \quad \text{للسحب مع الإعادة} \qquad (13-3)$$

$$n = \frac{N Z^2 s^2}{N d^2 + Z^2 s^2} = \left(\frac{N Z^2 r(1-r)}{N d^2 + Z^2 r(1-r)}\right) \quad \text{للسحب بدون اعادة} \qquad (14-3)$$

حيث أن : S^2 هو تباين العينة أو تقديره من أي عينة سابقة .

وأن: d هو مقدار الدقة المطلوبة وتحدد من قبل الجهات المعينة أومن قبل الباحث .

وأن: Z هي قيمة المتحول الطبيعي المعياري المقابل لنصف مستوى الدلالة $\frac{\propto}{2}$ على الطرفين.

وأن: r هو مقدار النسبة في المجتمع أو اي تقدير لها من خلال أي عينة اختبارية.

وفي حالة اختبار النسب المتوازنة في المجتمع نضع ($r = 0.50$) حتى نحصل على أكبر حجم ممكن للعينة، أما عندما يكون حجم المجتمع N كبيراً أو غير معروف، يفضل استخدام العلاقة (13−3) للسحب مع الإعادة , ثم نقوم بسحب العينة المذكورة من المجتمع ونحسب متوسطها $\bar{x}$ وتباينها S^2 من العلاقتين:

$$\bar{x} = \frac{1}{n}\sum x_i \qquad (15-3)$$

$$S^2 = \frac{1}{n-1} * \sum (x_i - \bar{x})^2 \qquad (16-3)$$

4−نقوم بحساب مؤشر الاختبار للمتوسط (أو للنسبة ضمن القوسين) من العلاقة المعيارية التالية:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \left(\frac{|r - R_0|}{\sqrt{\frac{R_0(1-R_0)}{n}}}\right) \quad : \quad \left(\sigma \text{ معلوم}\right) \qquad (17-3)$$

وهو متحول عشوائي يخضع للتوزيع الطبيعي المعياري $N(0,1)$، ولكن بما أن تباين المجتمع σ^2 وبالتالي انحرافه المعياري σ يكون غالباً مجهولاً، فإن حساب قيمة Z السابقة يكون أمراً مستحيلاً . ولكي نتخلص من هذه المشكلة نستبدل σ^2 بتباين العينة S^2 كتقدير جيد له، ونعرف مؤشر جديد لاختبار متوسط المجتمع μ (دون تعديل المقام في مؤشر النسبة لأن R_0 تكون معلومة من فرضية العدم H_0) . بالعلاقة التالية :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \left(\frac{|r - R_0| - \frac{1}{2n}}{\sqrt{\frac{R_0(1 - R_0)}{n}}} \right) \qquad (18 - 3)$$

وهو متحول جديد يخضع لتوزيع (ستودينت) بـ $(n - 1)$ درجة حرية عند اختبار المتوسط, وللتوزيع الطبيعي عند اختبار النسبة .

ملاحظة: إذا كان حجم العينة n كبيراً $(n > 30)$، فإن توزيع (ستودينت) يقترب من التوزيع الطبيعي المعياري، وعندها نعتبر t في العلاقة (3–18) خاضعاً للتوزيع الطبيعي المعياري .

5–نقوم بتحديد منطقتي الرفض والقبول واتخاذ القرار:

لتحديد منطقتي الرفض والقبول ولاتخاذ القرار اللازم حول صحة الفرضية H_0، يوجد طريقتان لاتخاذ القرار المناسب هما: طريقة القيمة الحرجة . وطريقة احتمال الدلالة P . وسنشرحهما كما يلي :

أ– **طريقة القيمة الحرجة** لـ Z أو t : لنفترض أن الاختبار ثنائي الجانب (أي أن $H_1: \mu \neq \mu_0$)، فعندها يكون مستوى الدلالة ∝ موزعاً على الجانبين، وهنا يكون لدينا حالتان لـ σ هما: إما أن تكون قيمة σ معلومة، أو أن تكون σ مقدرة من العينة بـ S، ولذلك نعالجها كما يلي :

– إذا كانت قيمة σ معلومة: فإننا نحسب قيمة مؤشر الاختبار من العلاقة (3–17) ثم نقارنها مع القيمة الحرجة لمتحول التوزيع الطبيعي $Z_{\frac{\alpha}{2}}$، لذلك نبحث في جدول التوزيع الطبيعي المعياري عن القيمة الحرجة $Z_{\frac{\alpha}{2}}$ المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$ على الطرفين، ثم نقارن القيمة المحسوبة Z معها، ونتخذ القرار حول H_0 كما يلي:

إذا كانت $|Z| \leq Z_{\frac{\alpha}{2}}$، تكون Z واقعة في منطقة القبول, وعندها نقبل فرضية العدم ونقول بأن $\mu = \mu_0$ ، باحتمال ثقة يساوي ∝−1 .

إذا كانت $|Z| > Z_{\frac{\alpha}{2}}$، تكون Z واقعة في منطقة الرفض وعندها نرفض فرضية العدم ونقول بأن$\mu \neq \mu_0$، بمستوى دلالة يساوي ∝ .

– أما عندما يكون σ^2 مجهولاً, فإننا سنحسب قيمة المؤشر t من العلاقة (3–18) , ثم نبحث في جدول توزيع (ستودينت) عن القيمة الحرجة $t_{\frac{\alpha}{2}}$ المقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$ على الطرفين ولدرجة حرية $(n - 1)$ ونتخذ القرار بالمقارنة كما يلي:

إذا كانت $|t| \leq t_{\frac{\alpha}{2}}$ نقبل فرضية العدم ونقول بأن $\mu = \mu_0$ باحتمال ثقة ∝−1

إذا كانت $|t| > t_{\frac{\alpha}{2}}$ نرفض فرضية العدم ونقول بأن $\mu \neq \mu_0$ بمستوى دلالة ∝

ملاحظة: إذا كان الاختبار أحادي الجانب (يميني ويساري) , فإن القيمة الحرجة لـ Z هي القيمة المقابلة لكامل مستوى الدلالة ∝ ونرمز لها بـ Z_{α}، وعندها نتخذ القرار عند المقارنة حول H_0 كما يلي:

- إذا كان الاختبار أحادي يميني $(H_1: \mu > \mu_0)$ فإننا نقارن قيمة Z المحسوبة مع $Z_\propto$, فإذا كانت $Z \le Z_\propto$ نقبل فرضية العدم H_0، ونقبل بأن $\mu = \mu_0$، أما إذا كانت $Z > Z_\propto$ فإننا نرفض H_0 ونعترف بأن $\mu > \mu_0$ كما في الشكل (3−5) السابق .
- أما إذا كان الاختبار أحادي يساري $(H_1: \mu < \mu_0)$ فإننا نقارن قيمة Z المحسوبة مع $(-Z_\propto)$ السالبة . فإذا كانت $(Z \ge -Z_\propto)$ نقبل فرضية العدم H_0 ونقبل بأن $\mu = \mu_0$، أما إذا كانت $(Z < -Z_\propto)$ فإننا نرفض H_0 ونقبل H_1 ونعترف بأن $\mu < \mu_0$ كما في الشكل (1−6) السابق.

وكذلك الأمر عند استخدامنا لمؤشر (ستودينت) t فإن القيمة الحرجة لمتحوله t هي القيمة المقابلة لكامل مستوى الدلالة ∝ ولدرجة الحرية $(n-1)$ ونرمز لها بـ $t_\propto$. ونتخذ القرار كما يلي:

- إذا كان الاختبار أحادي يميني $(H_1: \mu > \mu_0)$ وكان $t \le t_\propto$ فإننا نقبل فرضية العدم H_0 ونقبل بأن $\mu = \mu_0$، أما إذا كان $t > t_\propto$ فإننا نرفض H_0 ونقبل H_1 ونعترف بأن $\mu > \mu_0$.
- أما إذا كان الاختبار أحادي يساري $(H_1: \mu < \mu_0)$ فإننا نقارن قيمة t المحسوبة مع $(-t_\propto)$ السالبة، فإذا كانت $(t \ge -t_\propto)$ فإننا نقبل الفرضية H_0 ونقبل بأن $\mu = \mu_0$، أما إذا كانت $(t < -t_\propto)$ فإننا نرفض H_0 ونقبل H_1 التي نقول بأن: $\mu < \mu_0$.

مثال (3−1): لنفترض إننا نريد اختبار أن يكون توقع X في المجتمع $\mu_0 = 50$، فسحبنا عينة عشوائية منه بحجم $n = 25$ عنصراً . فكان متوسطها: $\bar{x} = 53$ وتباينها: $S^2 = 400$, ثم حددنا مستوى الدلالة بـ $\propto = 0.05$ ووضعنا الفرضيتين كما يلي:

$$H_0: \mu = 50 \qquad H_1: \mu \neq 50 \qquad \text{(الاختبار ثنائي الجانب)}$$

ولإجراء هذا الاختبار نلاحظ أن تباين المجتمع σ^2 مجهول. لذلك نستخدم العلاقة (3−18) والتي تأخذ الشكل التالي :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{53 - 50}{20/5} = \frac{3}{4} = 0.75$$

وبما أن t يخضع لتوزيع (ستودينت) بـ $(n-1)$ درجة حرية , فإننا نقوم بحساب القيمة الحرجة له $t_{n-1}\left(\frac{\propto}{2}\right)$ ، فنجد من جداول (ستودينت) أن:

$$t_{n-1}\left(\frac{\propto}{2}\right) = t_{24}\left(\frac{0.05}{2}\right) = t_{24}(0.025) = 2.064$$

وبمقارنة القيمة المحسوبة $t = 0.75$ مع القيمة الحرجة $t_{24}\left(\frac{\propto}{2}\right) = 2.064$ نجد أن $|t| < t_{24}\left(\frac{\propto}{2}\right)$ لذلك نقبل فرضية العدم H_0 والتي تقول أن $\mu = 50$, وذلك باحتمال ثقة $1 - \propto = 0.95$.

ملاحظة: إذا كان تباين المجتمع σ^2 معلوماً نطبق العلاقة (3−7) ونستخدم المتحول المعياري Z .

ب− **طريقة اتخاذ القرار حول H_0 باستخدام احتمال الدلالة P** (P- Value) أو (Sig)

لتوضيح كيفية تطبيق هذه الطريقة لابد من توضيح معنى احتمال الدلالة (Signification probability) والذي يرمز له في البرامج الحاسوبية بالرمز P أو(P-value) أو بالرمز (Sig). إن احتمال الدلالة P حسب التعريف هو: الاحتمال الذي تتركه القيمة المحسوبة لمؤشر الاختبار Z أو t أو غيرهما على طرفي التوزيع الاحتمالي، أو على أحد طرفيه، وهو يتأثر بنوع الاختبار وبحالاته المختلفة التالية:

فإذا كان الاختبار ثنائي الجانب $(H_1 , \mu \neq \mu_0)$، فإن قيمة الاحتمال P، يتم توزيعها بالتساوي على طرفي التوزيع، بحيث يكون لكل طرف $\frac{P}{2}$، والطرفان هنا يقابلان المجالين المفتوحين $]+Z , +\infty[$ و $]-\infty , -Z[$

أما إذا كان الاختبار أحادي يميني $(H_1 , \mu > \mu_0)$، فإن كامل قيمة P تكون متوضعة على اليمين وتقابل المجال المفتوح $]+Z , +\infty[$.

وإذا كان الاختبار أحادي يساري $(H_1 , \mu < \mu_0)$، فإن كان كامل قيمة P تكون متوضعة على اليسار وتقابل المجال المفتوح $]-\infty , -Z[$.

أي أن الاحتمال P يساوي المساحة التي تقع تحت منحني التوزيع وتقابل المجالات المذكورة (حسب كل حالة)، ويتم تحديد القيمة العددية لـ P بعد إعداد الحسابات اللازمة والقيام بإجراء الاختبار المفروض والحصول على القيمة المحسوبة Z أو t أو غيرهما . ثم حساب قيمة تكامل التوزيع الاحتمالي على المجالات المذكورة (حسب كل حالة)، كما سنرى لاحقاً .

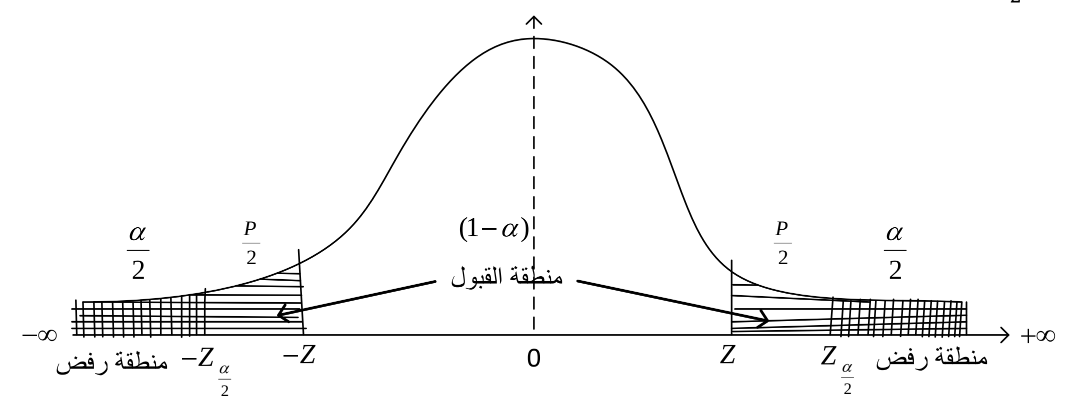
واخيراً نشير إلى أن احتمال الدلالة P يختلف جذرياً عن مستوى الدلالة $\propto$ (Signification Level), الذي يحدده الباحث أو المشرفون على البحث قبل إجراء البحث وقبل إجراء الاختبار نفسه .

ويستفاد من P في اتخاذ القرارات حول الفرضية H_0 وذلك بمقارنتها مع $\propto$.

ولتوضيح ذلك نأخذ حالة التوزيع الطبيعي المعياري، ونحسب منه قيمة احتمال الدلالة p، حسب حالات الاختبار التالية:

1) حالة الاختبار ثنائي الجانب: أي أن منطقة الرفض حسب القواعد السابقة تقع على الجانبين.

- فإذا كانت $|Z| \leq Z_{\frac{\propto}{2}}$ فإننا نقبل فرضية العدم H_0 ويكون لدينا الشكل التالي:



شكل (3-7) تحديد المساحة P

من هذا الشكل نلاحظ أن $\frac{\propto}{2}$ هي المساحة المظللة بخطوط عمودية وهي التي تقابل المجال $\left]Z_{\frac{\propto}{2}}, +\infty\right[$.

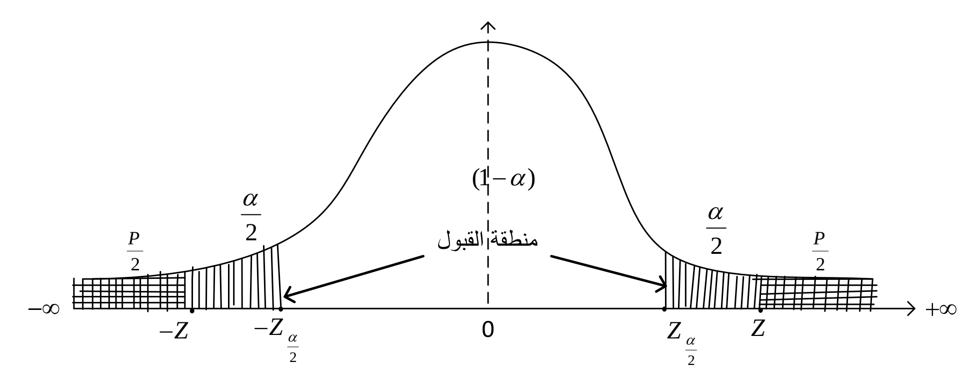
أما $\frac{P}{2}$ فهي المساحة المظللة بخطوط أفقية وهي التي تقابل المجال $]Z, +\infty[$، وعندما تكون $Z < Z_{\frac{\propto}{2}}$ كما في الشكل (1−7) يكون لدينا $\frac{P}{2} > \frac{\propto}{2}$ ، ويكون لدينا $P >\propto$، أي أنه علينا أن نقبل الفرضية H_0 إذا كانت $P >\propto$ (لأنه يكون لدينا $|Z| < Z_{\frac{\propto}{2}}$) .

ويتم حساب قيمة الاحتمال P في هذه الحالة من تكامل التوزيع الطبيعي المعياري على المجالين المتناظرين $]Z, +\infty[$ و $]-\infty, -Z[$ كما يلي:

$$P = \int_{-\infty}^{-Z} f(x)dx + \int_{Z}^{+\infty} f(x)dx = 2 * \int_{Z}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} * dx = 2[1 - \phi(Z)] \quad (19-3)$$

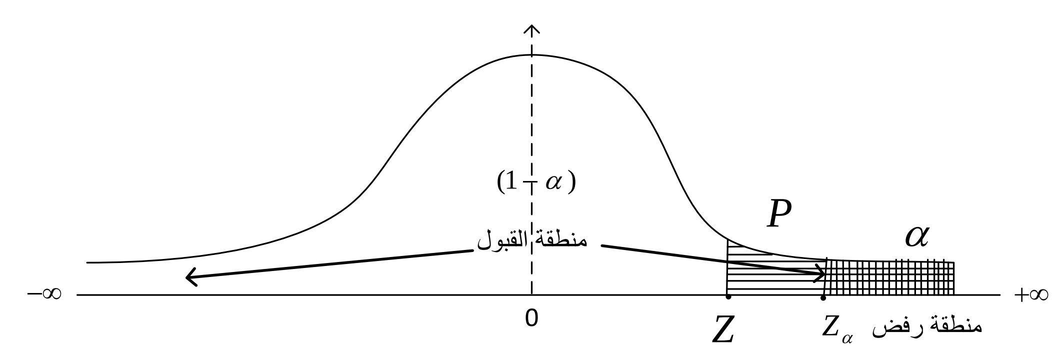
ويمكن الحصول على قيمة هذا التكامل من الجداول الإحصائية الجاهزة أو من الحواسيب المبرمجة .

– أما إذا كانت $|Z| > Z_{\frac{\propto}{2}}$ فإننا نرفض فرضية العدم H_0، وعندها يكون لدينا $\frac{P}{2} < \frac{\propto}{2}$ ، أي يكون لدينا $P <\propto$، أي أنه يجب علينا أن نرفض فرضية العدم H_0 إذا كان $P <\propto$ (لأن $|Z| > Z_{\frac{\propto}{2}}$) كما هو موضح في الشكل (1−8) التالي، حيث رمزنا للمساحة المظللة بخطوط أفقية من الطرفين للاحتمال P وللمساحة المظللة بخطوط عمودية لمستوى الدلالة $\propto$.



شكل (3−8) تحديد المساحة P

2) <u>حالة الاختبار الأحادي اليميني</u>: فإذا كانت $Z \leq Z_{\propto}$ فإننا نقبل فرضية العدم H_0، ويكون لدينا الشكل التالي:



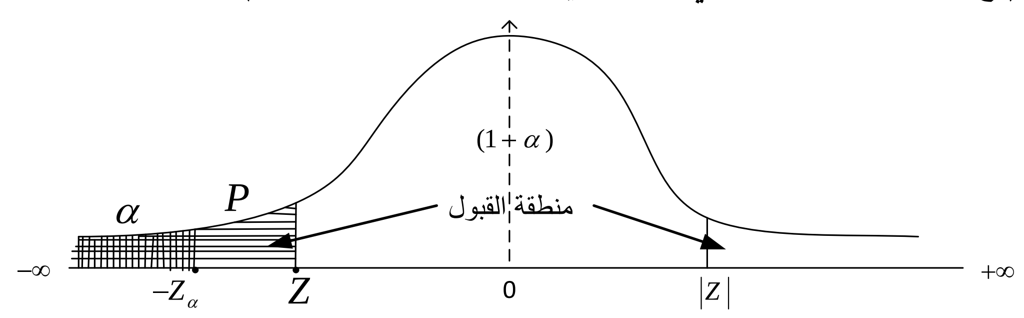
الشكل (3−9) تحديد المساحة P

ومن الشكل (الشكل (3−9) نلاحظ أن ∝ هي المساحة المظللة بخطوط عمودية وهي التي تقابل المجال]$Z_{\propto}$, +∞[، أما P فهي المساحة المظللة بخطوط أفقية وهي التي تقابل المجال]Z , +∞[، وهنا يكون لدينا $P >\propto$، أي أنه علينا أن نقبل فرضية العدم H_0 إذا كان $P >\propto$ (لأن يكون $Z \leq Z_{\propto}$)

− أما إذا كانت $Z > Z_{\propto}$ (تقع على يمينها) فإننا نرفض الفرضية H_0، وعندها يكون لدينا $P <\propto$، أي أنه علينا أن نرفض الفرضية H_0 إذا كان $P <\propto$ (لأن $Z > Z_{\propto}$)، ويتم حساب قيمة P في هذه الحالة من تكامل التوزيع الطبيعي المعياري على المجال]Z , +∞[كما يلي :

$$P = \int_{Z}^{+\infty} f(x)d_x = \int_{Z}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} * dx = 1 - \phi(Z) \qquad (20-3)$$

3) <u>حالة الاختبار الأحادي اليساري</u>: إذا كانت $Z > -Z_{\propto}$ فإننا نقبل فرضية العدم H_0 ويكون لدينا الشكل التالي: (مع ملاحظة أن قيمة Z هي قيمة جبرية فقد تكون سالبة أو موجبة)



الشكل (3− 10) تحديد المساحة P

ومن الشكل (3− 10) نلاحظ أن ∝ هي المساحة المظللة بخطوط عمودية وهي التي تقابل المجال]$-\infty, -Z_{\propto}$[، أما الاحتمال P فهو المساحة المظللة بخطوط أفقية وهي المساحة التي تقابل المجال]$-\infty$, Z[، وهذا يعني أن $P >\propto$، لذلك يجب علينا أن نقبل الفرضية H_0 إذا كان $P >\propto$ (لأن $Z > -Z_{\propto}$) . [انتبه إلى ذلك الاختلاف] .

− أما إذا كانت $Z < -Z_{\propto}$ (تقع على يسارها) فإننا نرفض الفرضية H_0، وعندها يكون لدينا$P <\propto$، أي أنه علينا أن نرفض الفرضية H_0 إذا كانت $P <\propto$ (لأن $Z < -Z_{\propto}$)، ويتم حساب P في هذه الحالة من العلاقة التالية:

$$P = \int_{-\infty}^{Z} f(x)d_x = \int_{|Z|}^{+\infty} f(x)d_x = \int_{|Z|}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} * dx = \phi(Z) \quad (21-3)$$

ملاحظة: إذا كان مؤشر الاختبار يخضع لتوزيع (ستودينت) t أو لأي توزيع آخر مثل X^2 أو F، فإن قيمة P تحسب حاسوبياً من تكاملات مشابهة للتكاملات السابقة على تلك التوزيعات وعلى المجالات المناسبة والمشابهة لتلك المجالات المذكورة. وهي أمور كثيرة وطويلة لا مجال للخوض فيها في هذا الفصل .

مثال (3−2) : لنفترض إننا نريد التأكد من نتيجة الاختبار في المثال (3−1), لذلك قمنا بسحب عينة أخرى كبيرة بحجم $n = 100$ عنصراً ثم حسبنا متوسطها وتباينها فكانا كما يلي $\bar{x} = 54$ و$S^2 = 225$, ثم حددنا مستوى الدلالة بـ (0.05 =∝) ووضعنا الفرضيتين كما يلي :

$$H_0: \mu = 50 \qquad H_1: \mu \neq 50$$ (الاختبار ثنائي الجانب)

وهنا نلاحظ أيضاً أن تباين المجتمع σ^2 مجهول . لذلك يجب علينا أن نطبق العلاقة (3−18)، ولكن بما أن حجم العينة n كبيراً $(n > 30)$ فإن تلك العلاقة تقترب من العلاقة (3−17) ويصبح t متقارباً مع Z ونكتبها كما يلي:

$$t \approx Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{54 - 50}{15/5} = \frac{4}{1.5} = 2.667$$

وبما أن Z يخضع للتوزيع الطبيعي المعياري، فإننا نقوم بإيجاد القيمة الحرجة $Z\left(\frac{\propto}{2}\right)$ من جداول التوزيع الطبيعي المعياري فنجد أن:

$$Z\left(\frac{\propto}{2}\right) = Z\left(\frac{0.05}{2}\right) = Z(0.025) = 1.96$$

وبمقارنة قيمة $Z = 2.667$ المحسوبة مع $Z\left(\frac{\propto}{2}\right) = 1.96$ الحرجة نجد أن:

$|Z| > Z\left(\frac{\propto}{2}\right)$. لذلك نرفض فرضية العدم H_0 التي تقول أن $\mu = 50$ ونقبل الفرضية H_1 التي تقول أن توقع X في ذلك المجتمع يختلف عن 50 . وذلك باحتمال ثقة 0.95 .

ملاحظة: يمكننا أن نستخدم طريقة احتمال الدلالة P لاتخاذ القرار حول H_0 ، لذلك نقوم بحساب P من العلاقة (3−19) فنجد من جداول التوزيع الطبيعي المعياري أن:

$$P = 2[1 - \phi(Z)] = 2[1 - \phi(2.667)] = 2[1 - 0.99615] = 0.0077$$

وبمقارنة P المحسوبة مع ∝ المفروضة نجد أن $P <\propto$. لذلك نرفض فرضية العدم H_0 بمستوى دلالة ∝ .

وهنا نلاحظ أن نتيجة الاختبار في هذا المثال تختلف عن المثال (3−1) وذلك لأن العينة مختلفة عن العينة الأولى بالبيانات والحجم .

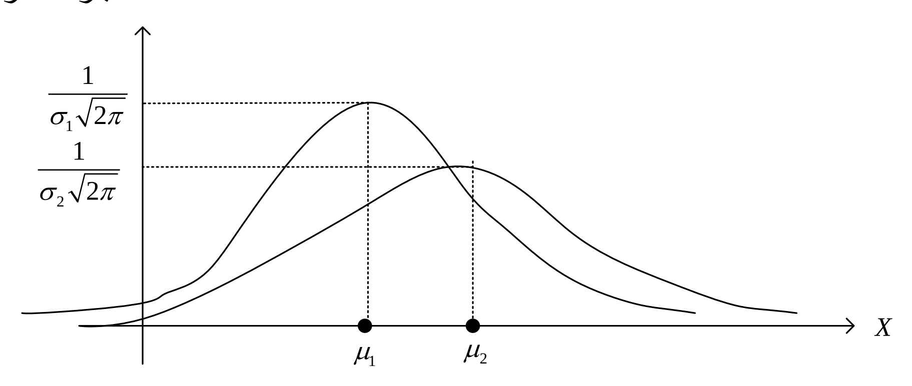
3-4: اختبارات معالم مجتمعين طبيعيين (من عينتين مستقلتين):

3−4−1: اختبار الفرق بين متوسطي مجتمعين طبيعيين:

لنفترض أننا نريد دراسة تغيرات متحول طبيعي X في مجتمعين منفصلين:

ولنفترض أن توقع X في المجتمع الأول هو μ_1 وتباينه فيه σ_1^2 وإن توزيعه الطبيعي هو $N(\mu_1 , \sigma_1^2)$, كما نفترض أن توقع X نفسه في المجتمع الثاني هو μ_2 وتباينه فيه σ_2^2، وإن توزيعه الطبيعي هو $N(\mu_2 , \sigma_2^2)$.

وبرسم هذين التوزيعين على شكل واحد نحصل على الشكل التالي :



الشكل (3−11) شكلان طبيعيان فيهما $\sigma_1^2 < \sigma_2^2$

ومن هذا الشكل نلاحظ أن عملية المقارنة بين μ_1 وμ_2 لا تتعلق بالفرق بينهما $(\mu_1 - \mu_2)$ فقط . بل تتعلق بشكل التوزيع الطبيعي وبتبايني X في هذين التوزيعين .

فإذا كان σ_1^2 و σ_2^2 مختلفان كثيراً فإن عملية المقارنة لا تكون متوازنة، لأنه إذا كان $\sigma_1^2 < \sigma_2^2$ فإن قيم X تكون متمركزة حول μ_1 أكثر من تمركز قيم X حول μ_2 .

وإن عملية المقارنة بين التوقعين μ_1 وμ_2 تكون أكثر فعالية عندما يكون $\sigma_1^2 = \sigma_2^2$ (أي عندما يكون الشكلان متشابهين) لذلك فإننا سنميز بين الحالتين التاليتين :

الحالة الأولى: الحالة التي يكون فيها $\sigma_1^2 \neq \sigma_2^2$

الحالة الثانية: الحالة التي يكون فيها $\sigma_1^2 = \sigma_2^2$

كما أننا سندرس الحالة التي يكون فيها σ_1^2 و σ_2^2 معلومين والحالة التي يكون فيها σ_1^2 و σ_2^2 مجهولين .

ولإجراء هذا الاختبار حول $(\mu_1 - \mu_2)$ علينا أن نسحب من هذين المجتمعين عينيتين عشوائيتين ومستقلتين بحجمين n_1 وn_2 ونحسب منهما مايلي:

أ- متوسط العينة الأولى $\bar{x}_1 = \frac{\sum x_i}{n_1}$ ويعتبر تقديراً غير متحيز لمتوسط المجتمع الأول μ_1 .

ب- متوسط العينة الثانية $\bar{x}_2 = \frac{\sum x_i}{n_2}$ ويعتبر تقديراً غير متحيز لمتوسط المجتمع الثاني μ_2 .

ج- نحسب التباين المصحح للعينة الأولى من العلاقة: $s_1^2 = \frac{1}{n_1-1}\sum(x_{i1} - \bar{x}_1)^2$,
ويعتبر هذا التباين تقديراً غير متحيز لتباين المجتمع الأول σ_1^2 .

د- نحسب التباين المصحح للعينة الثانية من العلاقة : $s_2^2 = \frac{1}{n_2-1}\sum(x_{i2} - \bar{x}_2)^2$، ويعتبر هذا التباين تقديراً غير متحيز لتباين المجتمع الثاني σ_2^2 .

هـ-نحسب الفرق بين متوسطين هاتين العينتين $(\bar{x}_1 - \bar{x}_2)$، وهو يعتبر تقديراً غير متحيز للفرق بين متوسطي المجتمعين$(\mu_1 - \mu_2)$, كما يعتبر الفرق $(\bar{x}_1 - \bar{x}_2)$ متحولاً عشوائياً جديداً يخضع للتوزيع الطبيعي الذي توقعه $(\mu_1 - \mu_2)$ وتباينه $Var(\bar{x}_1 - \bar{x}_2)$.

ولاختبار الفرق بين متوسطي المجتمعين $(\mu_1 - \mu_2)$ نضع الفرضيتين كما يلي :

فرضية العدم: $H_0: (\mu_1 - \mu_2) = 0$.

وتقابلها الفرضية البديلة والتي يمكن أن تكون على أحد الأشكال التالية:

على الشكل الثنائي الجانب : $H_1: (\mu_1 - \mu_2) \neq 0$.

أو على الشكل الأحادي اليميني: $H_1: (\mu_1 - \mu_2) > 0$.

أو على الشكل الأحادي اليساري: $H_1: (\mu_1 - \mu_2) < 0$.

ثم نقوم بتشكيل مؤشر الاختبار المعياري للفرق $(\bar{x}_1 - \bar{x}_2)$ من العلاقة التالية :

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{Var(\bar{x}_1 - \bar{x}_2)}} \qquad (19-3)$$

ثم نقوم بمعالجته حسب الحالات السابقة لـ σ_1^2 و σ_2^2 التالية:

الحالة الأولى: إذا كان $\sigma_1^2 \neq \sigma_2^2$ وكان σ_1^2 و σ_2^2 معلومين عددياً فإننا نجد أن تباين الفرق $(\bar{x}_1 - \bar{x}_2)$ لهاتين العينتين المستقلتين يساوي : [لعدم وجود ارتباط بين العينتين] .

$$Var(\bar{x}_1 - \bar{x}_2) = Var(\bar{x}_1) + Var(\bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \ (\text{معلوم})$$

وبالتعويض في (3-19) نحصل على مؤشر الاختبار الطبيعي المعياري التالي :

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad : \left(\sigma_1^2 \neq \sigma_2^2 \ \text{ومعلومين}\right) \qquad (20-3)$$

وهو متحول يخضع للتوزيع الطبيعي المعياري $N(0\,,1)$.

أما عندما يكون التباينان σ_1^2 و σ_2^2 مجهولين وغير متساويين، فإننا نستبدلهما بتقديريهما غير المتحيزين s_1^2 و s_2^2 والمحسوبين من العينتين فنحصل على مؤشر اختبار آخر t يخضع تقاربياً لتوزيع (ستودينت) وله درجة حرية معقدة (انظر المثال (3-1)) وهو يأخذ الشكل التالي :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad : \left(\sigma_1^2 \neq \sigma_2^2 \ \text{ومجهولين}\right) \qquad (21-3)$$

ملاحظة: يمكن استخدام مؤشر الاختبار الأخير t في اختبارات الفرق بين المتوسطين $(\mu_1 - \mu_2)$ إذا كان حجما العينتين n_1 و n_2 كبيرين، وعندها نعتبر درجة الحرية مساوية لأصغر العددين $(n_1 - 1)$ أو $(n_2 - 1)$ لأن قيمتها الحقيقية تكون قريبة منها .

الحالة الثانية: وهي الحالة التي يكون فيها تباينا المجتمعين متساويين ومعلومين, أي عندما يكون لدينا $\sigma_1^2 = \sigma_2^2 = \sigma^2$, حيث σ^2 هي القيمة المشتركة المعلومة لهما، وعندها تأخذ العلاقة (3- 20) الشكل التالي:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sigma * \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad : \left(\sigma^2 \ \text{معلوم}\right) \qquad (\mathbf{22-3})$$

وهو متحول يخضع للتوزيع الطبيعي المعياري $N(0\,,1)$.

أما عندما يكون التباين المشترك σ^2 مجهولاً , فإننا نقدره من خلال المتوسط الحسابي للتباينين S_1^2 وS_2^2 المصححين والمحسوبين من العينتين والمثقلين بـ (n_1-1) و (n_2-1) على الترتيب، فنحصل من العلاقة المركبة لهما على ما يسمى بالتباين المدمج pooled ونرمز له بالرمز S_P^2 ونكتبه كما يلي :

$$S_P^2=\widetilde{\sigma^2}=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-1)+(n_2-1)}=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} \qquad (23-3)$$

ويبرهن في الاحصاء الرياضي على أن التقدير S_P^2 هو تقدير غير متحيز للتباين المشترك σ^2، وبذلك تأخذ العلاقة (3-22) الشكل التالي:

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)_0}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad (24-3)$$

أو الشكل التالي:

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)_0}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad (25-3)$$

حيث استبدلنا في (3-22) التباين المشترك σ^2 بالتباين المدمج S_P^2 ، وهو مؤشر يخضع تقاربياً لتوزيع (ستودينت) بدرجة حرية (n_1+n_2-2)، ويمكن استخدامه في اختبارات الفروق بشرط أن يكون تباينا المجتمعين متساويين، لذلك يجب أن نتحقق أولاً من أن: $\sigma_1^2=\sigma_2^2=\sigma^2$ قبل تطبيق (3-25).

3-4-2: اختبار الفرق بين نسبتين في مجتمعين طبيعيين:

لاختبار الفرق بين نسبتي خاصتين R_1 وR_2 في مجتمعين طبيعيين، نفترض أولاً أن $\sigma_1^2\neq\sigma_2^2$ ونضع فرضية العدم كما يلي: $H_0: R_1-R_2=0$ ، والفرضية البديلة من الشكل : $H_1: R_1-R_2\neq 0$ أو غيره. وعندها يأخذ مؤشر الاختبار الأول (3-20) الشكل التالي:

$$Z=\frac{(r_1-r_2)-(R_1-R_2)_0}{\sqrt{\frac{R_1(1-R_1)}{n_1}+\frac{R_2(1-R_2)}{n_2}}} \qquad (26-3)$$

حيث أن: R_1 وR_2 معلومتان وأن r_1 وr_2 هما النسبتان في العينتين المسحوبتين.

أما عندما يكون $\sigma_1^2=\sigma_2^2=\sigma^2$، فإننا نقدر σ^2 من النسبة المتوسطة $\bar{r}$ ونحسب تقديره من العلاقة التالية:

$$\widetilde{\sigma^2}=\bar{r}(1-\bar{r}) \qquad (27-3)$$

حيث أن النسبة المتوسطة $\bar{r}$ تحسب من المتوسط المثقل للنسبتين r_1 وr_2 كما يلي:

$$\bar{r}=\frac{n_1r_1+n_2r_2}{n_1+n_2} \qquad (28-3)$$

وعندها تأخذ العلاقة (1- 26) شكلاً آخر هو التالي:

$$t=\frac{(r_1-r_2)-(R_1-\mathrm{R}_2)_0}{\sqrt{\frac{\bar{r}(1-\bar{r})}{n_1}+\frac{\bar{r}(1-\bar{r})}{n_2}}}=\frac{(r_1-r_2)-(R_1-R_2)_0}{\sqrt{\bar{r}(1-\bar{r})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad (29-3)$$

ومنها نحصل على المؤشر t الخاضع تقاربياً لتوزيع (ستودينت) بـ $(n_1 + n_2 - 2)$ درجة حرية, ولكنه عندما يكون $(n_1 + n_2 - 2) > 30$، فإنه يخضع تقاربياً للتوزيع الطبيعي المعياري. ولإجراء الاختبار نحسب قيمة المؤشر t من العلاقة (3−29)، ثم نقارنها مع قيمة $t_{\frac{\alpha}{2}}$ الحرجة والمقابلة لنصف مستوى الدلالة $\frac{\alpha}{2}$ من الطرفين ولدرجة الحرية $(n_1 + n_2 - 2)$، وبناء على هذه المقارنة نقبل أو نرفض فرضية العدم H_0 وفق القواعد المذكورة سابقاً .

مثال (3−3): لدراسة حالة الفروقات بين كميتي البروتين ACTB-1 عند مرضى السرطان ومرضى الربو مقارنة مع الأشخاص الطبيعيين, أجريت التجارب اللازمة على ثلاث عينات واستخلصت منها النتائج والبيانات التالية [من نتائج تجارب رسلان في ألمانيا عام 2018]:

المؤشر / العينة	حجم العينة n_i	متوسط كمية البروتين في العينة $\bar{x}_i$	الانحراف المعياري SD_i
مرضى الربو	25	1063.126	669.1437
الأشخاص الطبيعيين	42	1535.488	479.3964
مرضى السرطان	14	2350.761	1116.602

والمطلوب: اختبار الفرق بين متوسط البروتين عند مرضى السرطان ومتوسطه عند الأشخاص الطبيعيين. ثم اختبار الفرق بين متوسطه عند مرضى الربو ومتوسطه عند الأشخاص الطبيعيين، وذلك بمستوى دلالة $\alpha = 0.05$.

الحل: لاختبار الفرق بين متوسطي البروتين ACTB-1 عند مرضى السرطان μ_1 وعند الأشخاص الطبيعيين μ_2، نضع الفرضيتين (العدم والبديلة) كما يلي:

فرضية العدم بين متوسطي المجتمعين وتشير إلى أنه: $H_0: \mu_1 - \mu_2 = 0$

لا يوجد فرق بين المتوسطين μ_1 و μ_2

الفرضية البديلة: وتشير إلى أنه يوجد فروق بينهما $H_1: \mu_1 \neq \mu_2$

الحالة الأولى: وهي الحالة التي يكون فيها $\sigma_1^2 \neq \sigma_2^2$، وهي الحالة التي تشير إليها بيانات الجدول .

وفي هذه الحالة نقوم بحساب قيمة مؤشر الاختبار t من العلاقة العامة:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(2350.761 - 1535.488) - 0}{\sqrt{\frac{(1116.602)^2}{14} + \frac{(479.3964)^2}{42}}}$$

$$t = \frac{815.273}{\sqrt{89057.1447 + 5471.926388}} = \frac{815.273}{307.4558} = 2.65167$$

ولاتخاذ القرار حول H_0 نستخدم كلا الطريقتين التاليتين:

- طريقة القيمة الحرجة $t_{\frac{\alpha}{2}}$

ولاتخاذ القرار المناسب بطريقة القيمة الحرجة $t_{\frac{\propto}{2}}$ حول الفرضية H_0 نبحث في الجداول الإحصائية لتوزيع (ستودينت) عن القيمة الحرجة $t_{\frac{\propto}{2}}$ المقابلة لدرجة حرية مساوية لأصغر العددين: $(n_1 - 1)$ أو $(n_2 - 1)$ فنجد أن $13 = (1 - 14) = (n_1 - 1) = d\!f$، وبما أن الاختبار ثنائي الجانب, لأن $H_1: \mu_1 \neq \mu_2$ ، نجد أن القيمة الحرجة لـ t تساوي: $t_{\frac{\propto}{2},13} = t_{0.25\,,13} = 2.1604$

وبمقارنة t المحسوبة مع $t_{\frac{\propto}{2}}$ الحرجة نجد أن : $2.65167 > 2.1604$ أي أن $t > t_{\frac{\propto}{2}}$.

لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 التي تقول أن $\mu_1 \neq \mu_2$. أي أن متوسط البروتين عند مرضى السرطان لا يساوي متوسطه عند الأشخاص الطبيعيين باحتمال ثقة(0.95) على الأقل .

ملاحظة: كان يمكننا الاستفادة من بيانات العينة ووضع الفرضية البديلة H_1 على الشكل الأحادي اليميني $H_1: \mu_1 > \mu_2$, وفي هذه الحالة يجب علينا أن نقارن t المحسوبة مع القيمة الحرجة $t_{13}(\propto)$ المقابلة لجانب واحد . لذلك نبحث في جداول (ستودينت) عن القيمة الحرجة $t_{13}(\propto)$ فنجد أن :

$$t_{13}(\propto) = t_{13}(0.05) = 1.7709$$

وبالمقارنة نجد أن: $t = 2.65167 > 1.7709$ لذلك نرفض فرضية العدم أيضاً . ونقبل بأن $\mu_1 > \mu_2$, أي نقبل بأن متوسط البروتين عند مرضى السرطان أكبر من متوسطه عند الأشخاص الطبيعيين باحتمال ثقة(0.95) على الأقل.

- <u>طريقة الاحتمال P:</u>

لاتخاذ القرار المناسب بطريقة احتمال الدلالة P، علينا أن نقوم بحساب قيمة P المقابلة للقيمة المحسوبة $t = 2.65167$, وبما أن الاختبار ثنائي فهي تساوي ضعف المساحة المحسوبة من تكامل توزيع (ستودينت) على المجال $]t\,,+\infty[$. وهذا يقتضي تحديد درجة الحرية الدقيقة المعرفة في توزيع (ستودينت) المستخدم في هذا الاختبار، وهناك عدة طرق لحساب الدرجة $d\!f$ وأهمها الطريقتين التاليتين:

<u>الطريقة الدقيقة</u> لحساب P: وهي طريقة معقدة وتطبق في البرامج الحاسوبية، ولحساب درجة الحرية اللازمة تستخدم العلاقة الآتية [Triola, P.390] :

$$d\!f = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = 14.62907983 \qquad (30 - 3)$$

وهناك علاقة أخرى معدلة لـ (30−3) لحساب درجة الحرية في هذه الحالة وهي($\text{df}= d\!f - 2$).

وبما أن قيمة $d\!f$ التي حصلنا عليها كانت على شكل عدد كسري (غير صحيح)، لذلك نقوم بحساب قيمتي P المقابلتين لدرجتي الحرية الصحيحتين المجاورتين للقيمة $d\!f$ وهما 14 و15، فنجد من تكامل توزيع (ستودينت) على المجال $]2.65167\,,+\infty[$ أن :

$$P_{14} = 2 * (0.0094832) = 0.0189664 \qquad \text{(للجانبين)}$$

$$P_{15} = 2 * (0.0090654) = 0.0181308 \qquad \text{(للجانبين)}$$

ولحساب القيمة الحقيقية لـP المقابلة لدرجة الحرية الكسرية (14.629) نستخدم العلاقة التناسبية التالية:

$$P = P_{14} + (P_{15} - P_{14})(df - 14)$$
$$P = 0.0189664 + (-0.0008356)(0.6290783)$$
$$P = 0.01844074$$

وهي قيمة قريبة جداً من قيمة P التي نحصل عليها من الحاسوب، وهذا يعني أن الحاسوب يتبع الطريقة الدقيقة والمعقدة في حساب P، وبما أن $P < \propto$ فإننا نرفض فرضية العدم H_0 التي تقول أن $\mu_1 = \mu_2$ ونقبل الفرضية البديلة التي تقوم أن $\mu_1 \neq \mu_2$، أي أن متوسط البروتين عند مرضى السرطان لا يساوي متوسطه الطبيعي باحتمال ثقة أكبر(بكثير من)0.95 ، وهو يساوي:

$$1 - P = 1 - 0.01844 = 0.98156$$

- <u>الطريقة التقريبية لحساب P</u>:

لحساب قيمة P التقريبية المقابلة لـ $(t = 2.65167)$ في توزيع (ستودينت) نقوم بتحديد درجة الحرية df من أصغر العددين $(n_1 - 1)$ و$(n_2 - 1)$، فنجد أنها $(df = 14 - 1 = 13)$، ;كما يمكن تقديرها من نتيجة الطريقة الدقيقة السابقة كما يلي: $= 14.62907983 - 2 = 13\ df$.

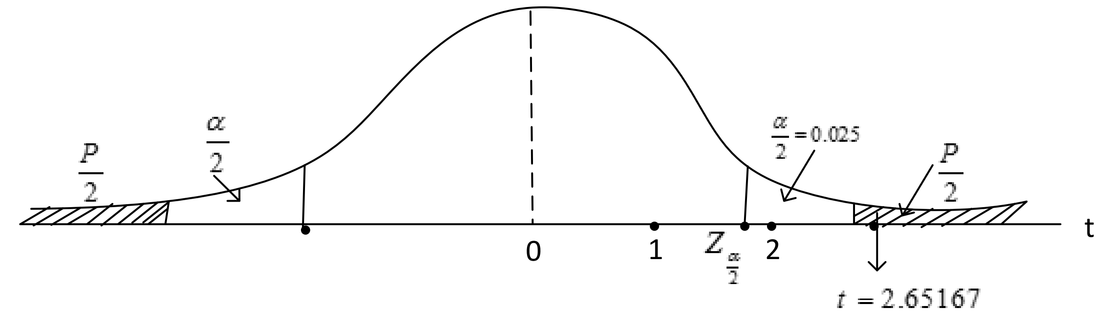
ومن الجداول الإحصائية لقيم متحول (ستودينت)، نجد أن قيمة $(t = 2.65167)$ المحسوبة والموافقة لدرجة حرية $df = 13$ تجعل الاحتمال P (من الطرفين) يساوي :

$$P = 0.0099741 * 2 = 0.0199482$$

وبما أن قيمة P أصغر من مستوى الدلالة $\propto = 0.05$, نرفض فرضية العدم H_0 التي تقول بعدم وجود فرق معنوي بين متوسطي المجتمعين، ونقبل الفرضية البديلة H_1 التي تقول أن متوسط المجتمع الأول (مرضى السرطان) لا يساوي المتوسط الطبيعي. وإن ذلك موثوق باحتمال ثقة أكبر بكثير من (0.95) , وهو يقترب من الواحد لأنه يساوي :

$$1 - P = 1 - 0.0199482 = 0.9800518 \approx 98\%$$

والشكل التالي يوضح معنى P بالمنطقة المظللة تحت منحنى التوزيع الطبيعي المعياري (المقارب لتوزيع (ستودينت)) :



الشكل (3-12) **تحديد المنطقة** P

الحالة الثانية: وهي الحالة التي يكون فيها: $\sigma_1^2 = \sigma_2^2 = \sigma^2$، وعندها نقوم بحساب مؤشر الاختبار t من العلاقة (3- 24) التالية :

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)_0}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$t=\frac{(2350.761-1535.488)-0}{\sqrt{\frac{13(1116.602)^2+41(479.3664)^2}{14+42-2}\left(\frac{1}{14}+\frac{1}{42}\right)}}$$

$$t=\frac{815.273}{212.60}=3.8345$$

علماً بأن t يخضع تقاربياً لتوزيع (ستودينت) بدرجة حرية n_1+n_2-2، ولهذا فإننا نقوم بحساب قيمة P المقابلة لـ $(t=3.8345)$ ولدرجة حرية $(df=14+42-2=54)$, فنجد أن الحاسوب يعطينا أن:

$$P=2*(0.00016548)=0.00033096$$

وهي قيمة أصغر بكثير من مستوى الدلالة $\propto=0.05$، لذلك نرفض فرضية العدم H_0 ونقبل الفرضية H_1 التي تقول بوجود فرق معنوي بين المتوسطين، ولكن هذه النتيجة محفوفة بالخطأ لأن الشرط المستخدم فيها $(\sigma_1^2=\sigma_2^2=\sigma^2)$ غير محقق في بيانات المثال المذكور، ولا يجوز الاعتماد عليها قبل إجراء اختبار لتساوي التباينين σ_2^2 وσ_1^2، ولقد قمنا بتطبيقها هنا للتدريب فقط .

ولاختبار الفرق بين متوسطي البروتين عند مرضى الربو والأشخاص الطبيعيين نتبع نفس الخطوات ونستخدم نفس العلاقات ونترك ذلك للقارئ على سبيل التدريب .

3-4-3: اختبار F لتساوي تبايني مجتمعين طبيعيين σ_2^2 وσ_1^2 :

لإجراء هذا الاختبار نضع فرضية العدم H_0 على الشكل التالي:

$$H_0=\frac{\sigma_1^2}{\sigma^2}=1 \qquad \Rightarrow(\sigma_1^2=\sigma_2^2) \qquad (31-3)$$

ونضع الفرضية البديلة كما يلي:

$$H_1=\frac{\sigma_1^2}{\sigma_2^2}>1 \qquad : \ (\sigma_1^2>\sigma_2^2) \qquad (32-3)$$

ثم نسحب من المجتمعين عينتين عشوائيتين بحجمين n_2و n_1، ونحسب تباينيهما المصححين S_2^2و S_1^2

ملاحظة: لتسهيل الحسابات تم تصميم جداول التوزيع F بحيث يكون رقم المجتمع الأول لصاحب التباين الأكبر (لذلك نرقم المجتمعين بحيث يكون $s_1^2 > s_2^2$ ، ونعدل الرموز في H_0 و H_1 حسب ذلك الترقيم) ثم نقوم بحساب مؤشر الاختبار F المعرف بالعلاقة التالية:

$$F=\frac{\frac{(n_1-1)s_1^2}{\sigma_1^2(n_1-1)}}{\frac{(n_2-1)s_2^2}{\sigma_2^2(n_2-1)}}=\frac{\frac{s_1^2}{\sigma_1}}{\frac{s_2^2}{\sigma_2^2}}=\frac{\sigma_2^2}{\sigma_1^2}*\frac{s_1^2}{s_2^2}\sim F_{v_1v_2} \qquad (33-3)$$

ولكن بما أن فرضية العدم تنص على أن: $\sigma_1^2=\sigma_2^2$ فإن المؤشر F يختصر ويأخذ الشكل التالي:

$$F = \frac{s_1^2}{s_2^2} \qquad : (s_1^2 > s_2^2) \qquad (34-3)$$

وهو متحول عشوائي يخضع لتوزيع فيشير $F_{(x)}$ بدرجتي حرية $v_1 = (n_1 - 1)$ للبسط و $v_2 =$ $(n_2 - 1)$ للمقام، وبعد حساب قيمة F نقارنها مع القيمة الحرجة $F_{(\propto)}$ (لاتجاه واحد) والمقابلة لكامل $\propto$ ولدرجتي الحرجة $v_1 = (n_1 - 1)$ و $v_2 = (n_2 - 1)$ ونتخذ القرار كما يلي :

إذا كانت $F \le F_{(\propto)}$ (أو كانت $P > \propto$) فإننا نقبل الفرضية H_0 التي تقول بتساوي التباينين $\sigma_1^2 = \sigma_2^2$. أما إذا كانت $F > F_{(\propto)}$ نرفض H_0 ونقبل H_1 ونقول بعدم تساوي التباينين المذكورين .

3-5: اختبارات معالم عدة مجتمعات طبيعية (من عدة عينات مستقلة):

3-5-1: اختبار تساوي متوسطات عدة مجتمعات طبيعية :

يطبق هذا الاختبار لمقارنة المتوسطات في أكثر من مجتمعين طبيعيين (3 فأكثر)، ولنفترض إننا سحبنا منهم عشوائياً عينات مستقلة بحجوم n_1 و n_2 و n_3 ... n_g (حيث أن g عدد المجتمعات و $g > 2$) . وكانت متوسطاتها $\bar{x}_1$ و $\bar{x}_2$ و $\bar{x}_3$ ... $\bar{x}_g$ وكان متوسط المتوسطات هو $\bar{\bar{x}}$.

وكانت تبايناتها المصححة s_1^2 و s_2^2 و s_3^2 ... s_g^2

فإننا نضع فرضيتي العدم والبديلة حول متوسطات هذه المجتمعات كما يلي:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_g \qquad (35-3)$$

$$H_1: \mu_k \neq \mu$$: من أجل k واحد على الأقل

كما نفترض أن تباينات هذه المجتمعات متساوية وتساوي σ^2 .

ثم نحسب مجاميع مربعات الانحرافات المختلفة وهي :

مجموع مربعات الانحرافات داخل العينات، أي مربعات (الخطأ) :

$$SSE = \sum_{k=1}^{g} (n_k - 1) s_k^2 \qquad (36-3)$$

حيث g: عدد المجتمعات .

مجموع مربعات الانحرافات بين العينات :

$$SSB = \sum_{k=1}^{g} n_i (\bar{x}_k - \bar{\bar{x}})^2 \qquad (37-3)$$

مجموع مربعات الانحرافات الكلية لجميع عناصر العينات :

$$SST = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{\bar{x}})^2 = \sum_{k=1}^{g} \sum_{i=1}^{n_k} x_{ki}^2 - \frac{T^2}{n} \qquad (38-3)$$

حيث أن: $n = \sum n_k$ وأن: $T = \sum_{k=1}^{g} \sum_{i=1} x_{ki}$

ويكون لدينا:

$$SST = SSB + SSE \qquad (39-3)$$

علماً بأن درجات الحرية لكل منهم، هي و$(g-1)$ و$(n-g)$ و$(n-1)$ على الترتيب، ثم نضع النتائج في جدول كالتالي :

جدول(3-3) : نتائج تحليل التباين الأحادي ANOVA

مصدر التباين	مجموع المربعات ورمزه	درجة الحرية	متوسط المربعات	قيمة F المحسوبة	قيمة F الحرجة	قيمة P
التباين بين العينات	SSB	$g-1$	$MSSB = \frac{SSB}{g-1}$	$F = \frac{MSSB}{MSSE}$	$F_{\propto}$	P
التباين داخل العينات	SSE	$n-g$	$MSSE = \frac{SSE}{n-g}$	ـــــــ	ـــــــ	ـــــــ
التباين الكلي	SST	$n-1$	ـــــــ	ـــــــ	ـــــــ	ـــــــ

ثم نحسب قيمة مؤشر الاختبار F المعرف بالعلاقة التالية :

$$F = \frac{MSSB}{MSSE} \qquad (40-3)$$

وهو يخضع للتوزيع F بدرجتي حرية $(g-1\ ,n-g)$ ونتعامل معه كما تعاملنا مع F السابقة عند اتخاذ القرار حول H_0 في اختبار تساوي التباينين . فإذا كان $F \le F(\propto)$ نقبل H_0 والعكس بالعكس .

ملاحظة: يسمى هذا الاختبار تحليل التباين ANOVA باتجاه واحد (Analysis Variance- one way) وسنقوم بدراسته بالتفصيل في الفصل السادس .

مثال (3-4): [مأخوذ من Copal P56 **بتصرف]**

لنفترض أنه لدينا (3) مجتمعات طبيعية، ونريد اختبار تساوي متوسطات متحول واحدX فيها, وبمستوى دلالة $\propto = 0.05$، لذلك سحبنا (3) عينات عشوائية منها بحجوم: $n_1 = 3 \quad n_2 = 5 \quad n_3 = 4$, ولنفترض أن البيانات الأصلية (غير الموجودة) أعطتنا أن مجاميع قياسات X فيها كانت تساوي ما يلي:

$$\sum_{i=1}^{3} x_{1i} = 53.5 \quad , \quad \sum_{i=1}^{5} x_{2i} = 102.5 \quad , \quad \sum_{i=1}^{4} x_{3i} = 64.4$$

وإن مجموعها الكلي يساوي:

$$T = \sum_{k=1}^{3}\sum_{i=1}^{n_k} x_{ki} = 53.5 + 102.5 + 64.4 = 220.4$$

وإن متوسطات X في العينات المسحوبة تساوي:

$$\bar{x}_1 = \frac{53.5}{3} = 17.83 \quad , \quad \bar{x}_2 = \frac{102.5}{5} = 20.50 \quad , \quad \bar{x}_3 = \frac{64.4}{4} = 16.10$$

وإن المتوسط العام لـX فيها (أو المتوسط المثقل للمتوسطات) يساوي:

$$\bar{x} = \frac{T}{\sum n_i} = \frac{220.4}{12} = 18.37$$

ثم نقوم بحساب الكسر $\frac{T^2}{n}$ فنجد أن:

$$\frac{T^2}{n} = \frac{(220.4)^2}{12} = 4048.01$$

ثم نقوم بحساب SST من العلاقة (1−38) فنجد من البيانات الأصلية أن:

$$SST = \sum_{k=1}^{3}\sum_{i=1}^{n_k} x_{ki}^2 - \frac{T^2}{n} = [(954.43 + 2105.13 + 1037.98) - 4048.01]$$

$$SST = 4097.54 - 4048.1 = 49.53$$

ثم نقوم بحساب SSB من العلاقة (1−37) فنجد أن:

$$SSB = \sum_{k=1}^{3} x_k(\bar{x}_k - \bar{x})^2 = 3(17.83 - 18.37)^2 + 5(20.50 - 18.37)^2 + 416.10 - 18.37^2$$

$$SSB = 44.17$$

ثم نقوم بحساب SSE من العلاقة:

$$SSE = SST - SSB = 49.53 - 44.17 = 5.36$$

ثم نقوم بوضع نتائج هذه الحسابات في جدول كالتالي :

جدول (3−4): جدول ANOVA

مصدر التباين	مجموع المربعات	درجة الحرية	متوسطات المربعات	
بين العينات SSB	$SSB = 44.17$	$g - 1 = 2$	$MSSB = 22.09$	$F = \frac{22.09}{0.556} = 37$
داخل العينات SSE	$SSE = 5.36$	$n - g = 9$	$MSSE = 0.556$	ـــــــ
التباين الاجمالي SST	$SST = 49.53$	$n - 1 = 11$	ـــــــ	ـــــــ

ثم نقوم بحساب قيمة المؤشر F من العلاقة:

$$F = \frac{\frac{SSB}{g-1}}{\frac{SSE}{n-g}} = \frac{\frac{44.17}{2}}{\frac{5.36}{9}} = \frac{22.09}{0.556} = 37$$

ولمقارنة قيمة F المحسوبة مع قيمتها الحرجة $F_{v_1,v_2}(\propto)$ المقابلة لمستوى الدلالة $(\propto = 0.05)$ ولدرجتي الحرية $v_1 = g - 1 = 2$, $v_2 = n - g = 9$ ، علينا أن نبحث عن قيمة $F_{v_1,v_2}(\propto)$ في جداول F فنجد أنها تساوي :

$$F_{v_1,v_2}(\propto) = F_{2,9}(0.05) = 4.24$$

ثم نقوم بمقارنة F المحسوبة مع $F_{2,9}(0.05)$ الحرجة، فنجد أن $F > F_{2,9}(0.05)$، لذلك نرفض فرضية العدم H_0، التي تقول أن $H_0: \mu_1 = \mu_2 = \mu_s$، ونقبل الفرضية H_1 التي تقول أن أحد متوسطات هذه المجتمعات (على الأقل) يختلف عن الأخرى. (ولعله المجتمع الثاني لأن $\bar{x}_2 = 20.50$).

3-5-2: اختبار تساوي تباينات عدة مجتمعات طبيعية (من عدة عينات مستقلة):

3-5-2-1: اختبار (بارتليت Bartlet) لتساوي تباينات عدة مجتمعات:

لنفترض إننا نريد دراسة تباينات متحول X في g مجتمعاً طبيعياً أو شبه طبيعي $(g > 2)$. لذلك سحبنا منها g عينة عشوائية بحجوم مختلفة أو متساوية نرمز لها بـ $n_1 , n_2 , \dots n_k \dots n_g$، ولمجموع حجومها بـ $n = \sum_{k=1}^{n} n_k$. ثم نقوم بحساب متوسطاتها ورمزنا لها بـ $\bar{x}_1 , \bar{x}_2 , \dots \bar{x}_k \dots \bar{x}_g$، وحساب تبايناتها ورمزنا لها بـ $S_1^2 , S_2^2 , \dots S_k^2 , \dots S_g^2$، والآن لنفترض أن تبايناتها X في هذه المجتمعات هي :

$$\sigma_1^2 , \sigma_2^2 , \dots \sigma_k^2 , \dots \sigma_g^2$$

ثم نضع الفرضيتين حول تساوي هذه التباينات كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \dots = \sigma_g^2 \qquad (41-3)$$

$H_1: \sigma_k^2 \neq \sigma_\ell^2$: من أجل زوج واحد (k , ℓ) على الأقل

أي أننا نفترض في H_0 أن تباينات X في هذه المجتمعات متساوية، مقابل الفرضية البديلة H_1 التي تعني أنها غير متساوية من أجل مجتمعين على الأقل .

ولاختبار هذه الفرضية قام Bartlitt باستخراج مؤشر خاص وعرفه بالعلاقة التالية:

$$BT = \frac{(n-g)\, ln\, S_p^2 - \sum_{k=1}^{g}(n_k - 1)\, ln\, s_k^2}{1 + \left[\frac{1}{3(g-1)}\right]\left[\sum_{k=1}^{g} \frac{1}{n_k - 1} - \frac{1}{n-g}\right]} \sim \chi_{g-1}^2 \qquad (42-3)$$

حيث أن: n هو حجم العينة الكلية $n = \sum_{k=1}^{n} n_k$، و g عدد المجتمعات .

حيث أن: S_k^2 هو تباين X في العينة k .

وأن S_p^2 هو التباين المدمج المحسوب من العلاقة :

$$S_p^2 = \frac{\sum (n_k - 1) s_k^2}{n-g} \qquad (43-3)$$

وبرهن على هذا المؤشر يخضع تقاربياً لتوزيع χ_{g-1}^2 بدرجة حرية $(g-1)$.

لذلك فإننا عند اتخاذ القرار حول الفرضية H_0 نقارن القيمة المحسوبة BT مع القيمة الحرجة $\chi_{g-1}^2(\propto)$ ونتخذ القرار عند مستوى دلالة $\propto$ كما يلي:

إذا كانت $BT \leq \chi_{g-1}^2(\propto)$ نقبل الفرضية H_0 والتي تنص على أن التباينات متساوية, (44-3)

أما إذا كان $BT > \chi_{g-1}^2(\propto)$ فإننا نرفض H_0 ونقبل الفرضية البديلة H_1، التي تنص على أن التباينات غير متساوية في مجتمعين على الأقل .

مثال (3-5): لنفترض أنه لدينا (5) خطوط لعصر الزيتون (معاصر) ونريد دراسة فيما إذا كانت تباينات الانتاج اليومي فيها متساوية أم مختلفة . لذلك وضعنا الفرضيتين كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$$

من أجل خطين على الأقل : $H_1: \sigma_k^2 \neq \sigma_\ell^2$

ثم قمنا بسحب (5) عينات طبقية من إنتاج هذه الخطوط خلال أربعة أيام $(n_k = 4)$ فحصلنا منها على البيانات التالية:

جدول (3-5): كميات الإنتاج اليومية (بالكغ) حسب الخطوط والأيام

	الخطوط / رقم اليوم	الخط A	الخط B	الخط C	الخط D	الخط E	المجموع
	1	250	310	250	340	250	
	2	260	330	230	270	240	
	3	230	280	220	300	270	
	4	270	360	260	320	290	
n_k	الأعداد	4	4	4	4	4	$n = 20$
$\bar{x}_k$	المتوسطات	252.5	320.0	240.0	307.5	262.5	1382.5
S_k^2	التباينات	291.667	1133.33	333.33	891.667	491.667	3141.662
$ln\, S_k^2$	لوغاريتمات التباينات	5.6756	7.0329	5.8091	6.7931	6.1978	31.5094

ولمتابعة الحل قمنا أولاً بحساب بعض الخصائص الاحصائية لتلك البيانات ووضعنا في أسفل الجدول السابق .

والآن نقوم بحساب الكميات التي تدخل في تعريف الاختبار BT فنجد أن التباين المدمج S_p^2 يساوي (انظر الجدول السابق) :

$$S_p^2 = \frac{\sum^g (n_k - 1) s_k^2}{n - g} = \frac{3(\sum s_k^2)}{20 - 5} = \frac{3 * (3141.667)}{15}$$

$$S_p^2 = \frac{9425}{15} = 628.333$$

كما نجد أن الحد الثاني في البسط يساوي :

$$\sum_{k=1}^{g} (n_k - 1)\, ln\, s_k^2 = 3\left(\sum_{k=1}^{g} ln\, S_k^2\right) = 3(31.5094) = 94.5292$$

ثم نقوم بحساب المقام فنجد أنه يساوي:

$$C = 1 + \left(\frac{1}{3(5-1)}\right)\left[\left(\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3}\right) - \left(\frac{1}{20-5}\right)\right]$$

$$C = 1 + \frac{1}{12}\left[\frac{5}{3} - \frac{1}{15}\right] = 1.1333$$

نعوض نتائج هذه الحسابات في معادلة المؤشر BT فنجد أن:

$$BT = \frac{(20-5)\, ln(628.333) - 945292}{1.1333} = \frac{2.1167}{1.1333} = 1.8678$$

ثم نقوم بإيجاد القيمة الحرجة لـ $\chi^2_{g-1}(\propto)$ عند مستوى الدلالة $\propto= 0.05$ ودرجة الحرية $g-1=5-1=4$ فنجد أن:

$$\chi^2_{g-1}(\propto) = \chi^2_4(0.05) = 9{,}488$$

وبمقارنة القيمة المحسوبة للمؤشر BT مع القيمة الحرجة $\chi^2_4(\propto)$ نجد أن $9.488 > 1.8678$ ، لذلك نقبل فرضية العدم H_0 التي نقول أن تباينات الانتاج على تلك الخطوط متساوية وباحتمال ثقة 0.95 .

ملاحظة: يمكننا أيضاً دراسة تساوي متوسطات الانتاج على هذه الخطوط وعندها يجب أن نستخدم الاختبار F وإجراء تحليل التباين ANOVA كما فعلنا في المثال (3– 4) السابق، ونترك ذلك للقارئ على سبيل التمرين.

2-2-5-3: اختبار (ليفيني Levene) لتساوي التباينات في عدة مجتمعات (من عدة عينات مستقلة).

يستخدم هذا الاختبار لدراسة تساوي أو تجانس تباينات متحول X في عدة مجتمعات طبيعية، وهو يقدم لنا خدمة جليلة عند تطبيق الكثير من الاختبارات الإحصائية، التي تفترض أن تباينات X في المجتمعات المدروسة متساوية، لأنه يساعدنا على التحقق من صحة تلك الافتراضات، ويعتبر هذا الاختبار بديلاً لاختبار Bartlitt، ولكنه أقل حساسية منه في الاعتماد على التوزيع الطبيعي .

فإذا كان لدينا شك قوي بأن البيانات المستخدمة ليست مسحوبة من مجتمع طبيعي (أو شبه طبيعي) فإنه يفضل استخدام اختبار Bartlitt ، لأنه يعطينا نتائج أفضل منه .

ولإجراء هذا الاختبار نفترض أننا نريد اختبار تساوي تباينات متحول طبيعي X في عدة مجتمعات (أو مجموعات)، ولنفترض أن عدد تلك المجتمعات $g(2 < g)$ وسحبنا منها g عينة عشوائية بحجوم مختلفة أو متساوية: $n_1, n_2, \dots n_k \dots n_g$ حيث $(n = \sum n_k)$، وحصلنا منها على متوسطاتها : $\bar{x}_1, \bar{x}_2, \dots \bar{x}_k \dots \bar{x}_g$ ، وعلى تبايناتها التالية: $S_1^2, S_2^2, \dots S_k^2, \dots S_g^2$. والآن لنفترض أن تباينات X في تلك المجتمعات هي: $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, \dots \sigma_g^2$.

وبناء على ذلك نصيغ الفرضيتين الاحصائيتين كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \dots = \sigma_g^2 \qquad (45-3)$$

$H_1: \sigma_k^2 \neq \sigma_\ell^2$: من أجل زوج واحد (k, ℓ) على الأقل

أما مؤشر الاختبار فيعرف حسب Levene بالعلاقة التالية:

$$W = \frac{(n-g)}{(g-1)} \frac{\sum_{k=1}^{g} (\bar{Z}_k - \bar{Z})^2}{\sum\sum (Z_{ki} - \bar{Z}_k)^2} \quad : \left(n = \sum n_k\right) \qquad (46-3)$$

حيث أن المتحول Z هو تحويل من المتحول X وفق إحدى العلاقات الثلاثة التالية:

$1-$ $\quad Z_{ki} = |x_{ki} - \bar{X}_k|$: حيث أن : $\bar{X}_k$ متوسط X في العينة K $\qquad (47-3)$

$2-$ $\quad Z_{ki} = |x_{ki} - X'_k|$: حيث أن : X'_k وسيط X في العينة K $\qquad (48-3)$

$3-: Z_{ki} = |x_{ki} - X''_k|$: حيث أن : X''_k هو المتوسط المرتب لـ 10% الأولى من قيم X $\qquad (49-3)$

أما متوسطات Z فتحسب كما يلي:

$$\bar{Z}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} Z_{ki}$$: K متوسط القيم Z_{ki} في العينة (50 − 3)

$$\bar{Z} = \frac{1}{g}\sum_{i=1}^{n_k} \bar{Z}_k = \frac{1}{n}\sum_{k=1}^{g}\sum_{i=1}^{n_k} Z_{ki}$$: المتوسط العام لـ Z_{ki} (47 − 3)

وهنا نشير إلى أن التعاريف الثلاثة لـ Z_{ki} تساعدنا في تحديد حصانة وقوة اختبار (ليفيني)، ويقصد بمصطلح الحصانة قدرة الاختبار على عدم إعطاء إشارة مزيفة عن عدم تساوي التباينات عندما تكون البيانات غير خاضعة للتوزيع الطبيعي وتكون التباينات فعلياً متساوية، ويقصد بالقوة قدره الاختبار على اكتشاف التباينات غير المتساوية عندما تكون التباينات فعلياً غير متساوية .

وأخيراً نشير إلى أن مؤشر الاختبار W يخضع لتوزيع F بدرجتي حرية $v_1 = g - 1$ و $v_2 = \text{n} - g$ (حيث أن: $n = \sum_{k=1}^{g} n_k$) .

ولاتخاذ القرار حول نتيجة الاختبار نقارن قيمة W المحسوبة من العلاقة (3−46) بالقيمة الحرجة $F_{v_1,v_2}(\propto)$ ونتخذ القرار كما يلي:

إذا كانت $W \le F_{v_1,v_2}(\propto)$ نقبل فرضية العدم H_0

أما إذا كانت $W > F_{v_1,v_2}(\propto)$ نرفض H_0 ونقبل الفرضية H_1 (51 − 3)

مثال (3−6): لنفترض أنه لدينا (10) مجموعات من الطلاب، ونريد اختبار تساوي تباينات أعمارهم في تلك المجموعات، فسحبنا من كل مجموعة عينة عشوائية بحجوم متساوية: $(n_k = 5)$ طلاب، فكان حجم العينة الكلية $n = 50$ طالباً . وبعد أخذ بيانات الأعمار في كل مجموعة وحساب متوسطاتها $\bar{x}_k$ وضعنا فرضيتي الاختبار كما يلي:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \dots = \sigma_{10}^2$$

$H_1: \sigma_k^2 \neq \sigma_\ell^2$: من أجل زوج واحد على الأقل

ثم نقوم بإجراء التحويلات من X إلى Z حسب إحدى العلاقات السابقة، ولتكن العلاقة (3−47) المستندة إلى المتوسطات $\bar{x}_k$، فنحصل على القيم Z_{ki}، ثم نقوم بحساب المتوسطات $\bar{Z}_k$ وأخيراً نقوم بحساب قيمة مؤشر الاختبار W ولنفترض أنها كانت تساوي:

$$W = \frac{(50-10)}{(10-1)}\frac{\sum 5(\bar{Z}_k - \bar{Z})^2}{\sum\sum(Z_{ki} - \bar{Z}_k)^2} = 1.75$$

وبما أن درجتي الحرية تساويان $v_1 = 10 - 1 = 9$ و $v_2 = 50 - 10 = 40$, فإننا نقارن القيمة المحسوبة W مع القيمة الحرجة $F_{v_1,v_2}(\propto)$. وباعتبار أن $\propto = 0.05$ نجد من جداول F أن:

$$F_{v_1,v_2}(\propto) = F_{9,40}(0.05) = 2.124$$

وبمقارنة القيمة المحسوبةW مع $F_{9,40}(0.05)$ نجد أن: $2.124 \le 1.75$ لذلك نقبل فرضية العدم H_0 التي تقول أن تباينات العمر X في هذه المجموعات متساوية وذلك باحتمال ثقة 0.95 على الأقل .

وبما أن التباينات متساوية فإننا نقوم بمقارنة الفروقات بين المتوسطات باستخدام العلاقة (3-40) أو (3-25) .

3-6: اختبار الأزواج المتقابلة (من عينتين مرتبطتين):

يطبق هذا الاختبار لمقارنة نتائج إجابات أو علامات عينية من نفس الأشخاص قبل التجربة وبعدها، لذلك تسمى الدرجات الأولى بالدرجات القبلية وتسمى الدرجات الثانية بالدرجات البعدية، ويمكن وضع النتائج في جدول خاص على شكل أزواج متقابلة (كل زوج لشخص واحد) فنحصل على عينتين مرتبطتين من الدرجات كما يلي:

جدول (3-6): البيانات المتقابلة

رقم الشخص	1	2	3	4	.....	i	.....	n	المتوسط	
الدرجات القبلية	x_1	x_2	x_3	x_4	.....	x_i	.....	x_n	$\bar{x}$	
الدرجات البعدية	y_1	y_2	y_3	y_4	.....	y_i	.....	y_n	$\bar{y}$	
الفروقات d_i	d_1	d_2	d_3	d_4	.....	d_i	.....	d_n	$\bar{d}$	S_d

ثم نقوم بحساب الفروقات بين قيمتي كل زوج من العلاقة : $d_i = x_i - y_{\text{i}}$

ثم نقوم بحساب متوسط وتباين هذه الفروقات من العلاقتين :

$$\bar{d} = \frac{1}{n}\sum d_i \qquad (52-3)$$

$$S_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2 = \frac{1}{n-1}\left[\sum d_i^2 - n\bar{d}^2\right] \qquad (53-3)$$

ولإجراء هذا الاختبار حول الفرق بين العينتين نضع الفرضيتين كما يلي: $H_0: \bar{d} = 0 \quad H_1: \bar{d} > 0$

حيث $\bar{D}$: هو متوسط الفروقات في المجتمع .

ثم نحسب الانحراف المعياري لـ d_i ونرمز له بـ S_d ، ثم نحسب مؤشر الاختبار المعرف بالعلاقة:

$$t = \frac{\bar{d} - \bar{d}_0}{S_d/\sqrt{n}} \qquad (54-3)$$

حيث $\bar{D}_0$ هي قيمة متوسط الفروقات المفترضة في المجتمع، وتؤخذ قيمتها الصفرية من فرضية العدم H_0، ثم نقارن قيمة t المحسوبة مع قيمة t الحرجة $t_{\propto,n-1}$ فإذا كانت $t < t_{\propto,n-1}$ تقبل فرضية العدم H_0، ونقول بأنه لا يوجد فرق بين الدرجات القبلية والبعدية، والعكس بالعكس .

مثال (3-7): لدراسة تأثير أحد الأدوية على مستوى ضغط الدم عند المرضى المصابين به. قرر أحد الباحثين إجراء تجربة هذا الدواء على (8) مرضى. ولذلك قام أولاً بقياس مستويات الضغط عند هؤلاء المرضى قبل إعطائهم الدواء . ثم قام بإعطائهم الدواء وبعد مرور ساعة على ذلك أخذ قياسات مستويات الضغط لهم، فحصل على البيانات القبلية والبعدية ثم قام بحساب الفروقات الزوجية D_i وقام بتربيعها فحصل على الجدول التالي :

جدول (3-8): بيانات المثال:

رقم المريض	1	2	3	4	5	6	7	8	المجموع
مستوى الضغط قبل التجربة X	170	175	180	175	160	170	175	180	___
مستوى الضغط بعد التجربة Y	150	160	170	160	170	160	170	185	___
الفروقات الزوجية d_i	20	15	10	15	−10	10	5	−5	+60
d_i^2	400	225	100	225	100	100	25	25	1200

ثم قام بحساب متوسط تلك الفروقات فوجد أن:

$$\bar{d} = \frac{\sum_{i=1}^{8} d_i}{n} = \frac{+60}{8} = 7.5$$

ثم قام بحساب تباين تلك الفروقات فحصل على أن :

$$S_d^2 = \frac{1}{n-1}\left[\sum d_i^2 - n\bar{d}^2\right] = \frac{1}{7}[1200 - 8(7.5)^2] = 107.14$$

$$S_d = \sqrt{S_d^2} = \sqrt{107.14} = 10.35$$

ثم قام بوضع فرضيتي الاختبار كما يلي:

$$H_0: \bar{d} = 0 \qquad H_1: \bar{d} > 0$$ (الاختبار أحادي يميني)

ثم قام بحساب قيمة مؤشر الاختبار من العلاقة :

$$t = \frac{\bar{d} - \bar{d}_0}{S/\sqrt{n}} = \frac{7.5 - 0}{10.35/\sqrt{8}} = 2.05$$

وباعتماد $\propto = 0.05$ قام بمقارنة هذه القيمة المحسوبة مع القيمة الحرجة $t_{n-1}(\propto)$ والتي تساوي: $t_{n-1}(\propto) = t_7(0.05) = 1.895$، فوجد أن: $t > t_7(0.05)$، لذلك رفض فرضية العدم H_0 التي تقول أن $(\bar{D} = 0)$ وتم قبول الفرضية البديلة H_1 التي تقول أن $(\bar{D} > 0)$. وهذا يعني أن متوسط الفروقات كان موجباً، وهو ما يؤكد أن الدواء المستخدم له تأثير إيجابي على مستويات الضغط عند المرضى .

تمارين الفصل الثالث

1- تقول شركة للمصابيح أن مصابيحها تعمل في المتوسط لمدة 1380 ساعة، ولاختبار ذلك أخذنا عينة بحجم $n = 100$ مصباحاً وراقبناها فوجدنا أن متوسط عملها كان $\bar{x} = 1300$ ساعة وأن تباينها كان $s^2 = 100$ اختبر، بمستوى دلالة $\alpha = 0,05$ ، صحة أدعاء الشركة .

2- من الدراسات السابقة وجدنا أن متوسط دخل العامل خلال العام الماضي كان 1500 ل.س وفي دراسة حالية شملت 150 عاملاً وجدنا أن متوسط دخلهم كان $\bar{x} = 1700$ وكان تباين العينة $s^2 = 2000$.
فهل يعد هذا دليلاً كافياً على تحسن دخل العمال؟ اختبر بمستوى دلالة قدره $\alpha = 0,05$.

3- **في استفتاء سريع حول أحد المرشحين وجدنا أن 300 شخص من أصل 500 شخص كانوا يؤيدون ذلك المرشح . فهل هذه النتائج تتناقض وادعاء المرشح بأن شعبيته تزيد على 80% من الناخبين؟ اختبر عندما** $\alpha = 0,05$.

4- لتكن $(x_1 , x_2 , \dots , x_n)$ عينة عشوائية من مجتمع طبيعي تباينه معلوم . ماهي أفضل منطقة رفض للفرضية $H_0 : \mu = 6$ مقابل الفرضية البديلة $H_1 : \mu = 4$ ؟

5- لدى تحليل عينتين من نوعين من التبغ بحجمين $n_1 = 5$ ، $n_2 = 6$ تجارب، كانت كمية النيكوتين مقدرة بالميلغرام كما يلي:

a	24	21	26	27	24	25
b	26	31	23	28	27	-

وبفرض أن نتائج التحليل خاضعة للتوزيع الطبيعي ولها تباين موحد . فهل تقدم هذه النتائج دليلاً كافياً على أن متوسط كمية النيكوتين في النوعين متساويتان ؟ اختبر بمستوى دلالة $\alpha = 0,05$.

6- **سحبنا عينة بحجم** $n = 100$ **من مجتمع طبيعي فوجدنا أن** $\bar{x} = 2,7$ و$s^2 = 23$ **فاختبر الفرضية** $H_0 : \mu = 3$ **بمستوى دلالة** $\alpha = 0,10$ **ثم اختبر الفرضية** $H_0 : \sigma^2 = 2,5$ **بمستوى دلالة** $\alpha = 0,10$.

7- سحبنا عينتين بحجمين $n_1 = n_2 = 11$ قياساً من مجتمعين طبيعيين متوسطاهما μ_1 وμ_2 على الترتيب ولهما تباين موحد . وحصلنا على $\bar{x}_2 = 60,4$, $\bar{x}_1 = 65,3$, $s_1^2 = 31,4$, $s_2^2 = 44,82$
فهل تقدم هذه المعلومات دلالة كافية على وجود فرق بين μ_1 وμ_2 ؟
اختبر بمستوى دلالة $\alpha = 0,10$ ثم اوجد مجال الثقة للفرق $(\mu_1 - \mu_2)$ ضمن مستوى الدلالة نفسه ؟

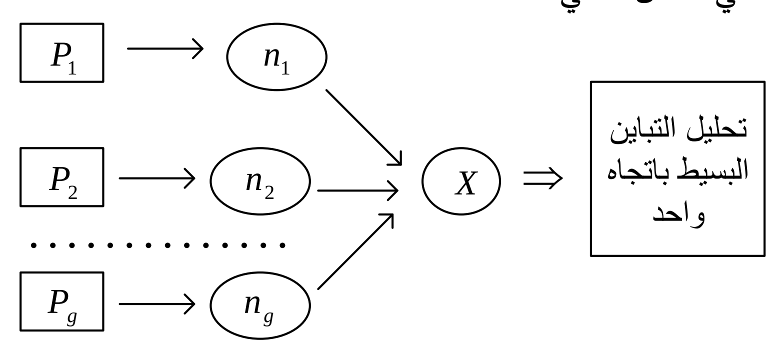
الفصل الرابع : تحليل التباين البسيط (ANOVA)

سنتناول في هذا الفصل عدة أنواع من تحليل التباين هي:

- تحليل التباين باتجاه واحد (ANOVA one way)
- تحليل التباين باتجاهين (ANOVA tow way)
- تحليل التباين بثلاث اتجاهات (ANOVA three way)
- تحليل المربع اللاتيني (LATIN SQUARE)
- تحليل التباين المشترك باتجاه واحد (ANCOVA one way)

4-1: تحليل التباين البسيط باتجاه واحد (ANOVA one way):

يتناول تحليل التباين البسيط باتجاه واحد دراسة تغيرات متحول واحد X (يسمى بالتابع), الناتجة عن عدة مجتمعات (أو معالجات) نرمز لها بـ $P_1\ P_2\ ...\ P_g$. وهنا يشترط أن يكون عدد المجتمعات $g > 2$ (لأنه إذا كان $g = 2$ فإننا نستخدم اختبار (ستودينت) t لمقارنة متوسطي المجتمعين), ويمكن تمثيل تأثير هذه المجتمعات على X كما في الشكل التالي:



الشكل (4-1): تمثيل ANOVA باتجاه واحد

ولدراسة تغيرات X الناتجة عن تأثيرات هذه المجتمعات نسحب من كل مجتمع k، عينة عشوائية بحجم n_k ، ثم نأخذ قياسات X من عناصر هذه العينات، ونضعها في جدول مناسب، يتضمن قياسات X من كل عينة n_k ومتوسطها $\bar{X}_k$ وتوقعها الرياضي في المجتمع μ_k , كما يمكن أن يتضمن تباينها S_k^2 كالجدول التالي :

جدول (4-1): قيم X حسب عينات المجتمعات

المجتمعات	حجوم العينات	قياسات X من عناصر العينات	متوسطات العينات	التوقعات في المجتمع	تباينات العينات
P_1	n_1	$x_{11}\ \ x_{12}\ \ x_{13}\ \ ...\ \ x_{1n_1}$	$\bar{X}_1$	μ_1	S_1^2
P_2	n_2	$x_{21}\ \ x_{22}\ \ x_{23}\ \ ...\ \ x_{2n_2}$	$\bar{X}_2$	μ_2	S_2^2
⋮	⋮	⋮	⋮	⋮	⋮
P_K	n_k	$x_{k1}\ \ x_{k2}\ \ x_{k3}\ \ ...\ \ x_{kn_k}$	$\bar{X}_K$	μ_k	S_K^2
⋮	⋮	⋮	⋮	⋮	⋮
P_g	n_g	$x_{g1}\ \ x_{g2}\ \ x_{g3}\ \ ...\ \ x_{gn_g}$	$\bar{X}_g$	μ_g	S_k^2

ويشترط في هذه العينات والبيانات أن تحقق الشروط أو الافتراضات التالية :

1- أن تكون العينات المسحوبة من المجتمعات عشوائية ومستقلة عن بعضها البعض .

2- أن يكون تباين X في جميع هذه المجتمعات موحداً ويساوي σ^2 .

3- أن تكون قيم X في كل مجتمع k خاضعة للتوزيع الطبيعي $N(\mu_K , \sigma^2)$. الذي توقعه μ_k وتباينه ثابت ويساوي σ^2 في كل المجتمعات .

ويمكننا تجاهل الشرط الثالث (حول الطبيعية) عندما تكون حجوم العينات كبيرة، والاستناد على مفعول نظرية النهاية المركزية في الاحتمالات، والتي تنص على أن توزيعات X في تلك المجتمعات تنتهي إلى التوزيع الطبيعي .

وعندها فإن فرضية العدم والفرضية البديلة حول توقعات هذه المجتمعات تأخذان الشكل التالي:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu_g \qquad (1-4)$$

$$H_1 : \mu_k \neq \mu_\ell$$

وذلك من أجل زوج واحد على الأقل $(K \neq \ell)$

وإذا رمزنا للمتوسط المثقل لهذه التوقعات بالرمز $\bar{\mu}$ والذي يحسب من العلاقة التالية :

$$\bar{\mu} = \frac{n_1\mu_1 + n_2\mu_2 + \cdots + n_g\mu_g}{n_1 + n_2 + \cdots + n_g} = \frac{\sum^g n_k\mu_k}{\sum_{k=1} n_k} = \frac{\sum^g n_k\mu_k}{n} \qquad (2-4)$$

حيث أن n_K هو حجم العينة المسحوبة من المجتمع k، وإن مجموعها = n، أي أن: $n = \sum_{K=1}^{g} n_k$

ويطلق على المتوسط $\bar{\mu}$ مصطلح المتوسط أو التوقع الكلي (Grand mean)، وبناء على ذلك يمكننا التعبير عن قيمة أي توقع μ_k بدلالة التوقع الكلي $\bar{\mu}$ كما يلي :

$$\mu_k = \bar{\mu} + (\mu_k - \bar{\mu}) \qquad (3-4)$$

أو على الشكل التالي :

$$\mu_k = \bar{\mu} + \tau_k \qquad (4-4)$$

حيث أن:

$$\tau_K = \mu_k - \bar{\mu} \qquad (5-4)$$

ونعبر عن ذلك لفظياً كما يلي:

(تأثير المجتمع k (المعالجة k)) + (التوقع الكلي) = (توقع X في المجتمع k)

وهذا يقودنا إلى تعديل فرضية العدم حول التوقعات μ_k إلى فرضية عدم جديدة H_0 مقابل H_1 كما يلي:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \cdots = \tau_g = 0 \qquad (6-4)$$

وذلك من أجل مجتمع واحد على الأقل :

$$H_1 : \tau_k \neq 0$$

وهذا يجعلنا نعتبر أن قياسات X في المجتمع k ، والتي سنرمز لها بـ x_{ki} ، خاضعة للتوزيع الطبيعي $N[(\bar{\mu} + \tau_k) , \sigma^2]$ وبالتالي يمكننا كتابة كل قياس منها x_{ki} كما يلي :

$$x_{ki} = \bar{\mu} + (\mu_k - \bar{\mu}) + (x_{ki} - \mu_k) = \bar{\mu} + \tau_k + \varepsilon_{ki} \qquad (7-4)$$

والتي يمكن كتابتها على الشكل التالي:

(حد الخطأ العشوائي) + (تأثير المجتمع k) + (التوقع الكلي) = (القياس x_{ki})

حيث أن: $\varepsilon_{ki} = x_{ki} - \mu_k$ ، وهو حد الخطأ العشوائي للقياس x_{ki} (أو البواقي)، وهي حدود مستقلة ويفترض أن تكون خاضعة للتوزيع الطبيعي $N(0\ ,\sigma^2)$. ولكن التأثيرات المجتمعية τ_k المعرفة في العلاقة (4−7) مرتبطة مع بعضها البعض , وذلك لأنه اعتماداً على العلاقة (4−2) نجد أن:

$$\sum_{k=1}^{g} n_k\tau_k = \sum_{k=1}^{g} n_k(\mu_k - \bar{\mu}) = \sum n_k\mu_k - \bar{\mu}\sum n_k = \sum n_k\mu_k - n\mu = 0$$

وبالتالي نحصل على أن:

$$\sum_{k=1}^{g} n_k * \tau_k = 0 \qquad (8-4)$$

وهذا يعني أن مجموع قيم τ_k المثقلة بأحجام العينات n_k يساوي الصفر. وبالتالي يكون متوسطها المثقل $\bar{\tau} = 0$

وبناء على التركيب (4−7), فإن تحليل التباين (ANOVA) في العينات يستخدم نموذجاً مشابهاً لـ (4−7)، للتعبير عن القياسات x_{ki} المشاهدة في العينات المسحوبة من تلك المجتمعات وذلك كما يلي:

$$x_{ki} = \bar{X} + (\bar{X}_k - \bar{X}) + (x_{ki} - \bar{X}_k) = \bar{X} + \tilde{\tau}_k + e_{ki} \qquad (9-4)$$

والذي يمكن كتابته في العينة على الشكل التالي:

(4−10) **(البواقي) + (تقدير تأثير المجتمع k) + (المتوسط الكلي في العينة) = (القياس x_{Ki} المشاهد)**

حيث أن $\bar{X}$ هو المتوسط الكلي في العينة ويحسب من $\bar{X} = \frac{1}{n}\sum n_k\bar{X}_k$ ويعتبر $\bar{X}$ تقديراً غير متحيز للتوقع الكلي في المجتمع $\bar{\mu}$.

وأن $\tilde{\tau}_k = (\bar{X}_k - \bar{X})$ هو تقدير لحد التأثير τ_k للمجتمع k، علماً بأن هذه الحدود يجب أن تحقق الشرط (4−8) التالي: $\sum n_k\tau_k = 0$.

وأن $e_{ki} = (\bar{x}_{ki} - \bar{X}_k)$ هو تقدير لحد الخطأ العشوائي ε_{ki} , ويسمى بحد البواقي(residual) في العينة k ، وهذه الحدود تشكل متحولات عشوائية, توقعاتها تساوي الصفر وتخضع لـ $N(0\ ,\sigma^2)$.

مثال: (4−1): لنفترض أننا نريد دراسة تغيرات أحد المتحولات X في 3 مجتمعات, فسحبنا منها ثلاث عينات بحجوم مختلفة هي: $n_1 = 3\ ,\ n_2 = 2\ ,\ n_3 = 3$. وبعد أخذ قياسات X من عناصر هذه العينات حصلنا على القياسات التالية :

$(n_1 = 3)$: المجتمع 1 : X_{1i} 9 , 6 , 9

$(n_2 = 2)$: المجتمع 2 : X_{2i} 0 , 2 ,

$(n_3 = 3)$: المجتمع 3 : X_{3i} 3 , 1 , 2

وعند حساب متوسطات X في هذه العينات نجد أنها تساوي ما يلي :

$$\bar{X}_1 = \frac{9+6+9}{3} = 8$$

$$\bar{X}_2 = \frac{0+2}{2} = 1$$

$$\bar{X}_3 = \frac{3+1+2}{3} = 2$$

وكذلك نجد أن المتوسط الكلي لها يساوي :

$$\bar{X} = \frac{1}{n}\sum n_k\bar{X}_k = \frac{3*8+2*1+3*2}{3+2+3} = 4$$

وللتحقق من صحة العلاقة (6–9) نحسب قيمتي القياسين x_{11} و x_{31} فنجد أن:

$$9 = x_{11} = \bar{X} + (\bar{X}_1 - \bar{X}) + (x_{11} - \bar{X}_1) = 4 + (8-4) + (9-8) = 4+4+1 = 9$$

………………………………………………………………………………………….

$$3 = x_{31} = \bar{X} + (\bar{X}_3 - \bar{X}) + (x_{31} - \bar{X}_3) = 4 + (2-4) + (3-2) = 4-2+1 = 3$$

وهكذا يتم حساب بقية القيم x_{ki} .

وإذا قمنا بتطبيق مثل تلك الحسابات على كل مشاهدة من المشاهدات السابقة، نحصل على التصفيفة (arrays) التالية (وليس المصفوفة) :

$$\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & - \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & - \\ 4 & 4 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & - \\ -2 & -2 & -2 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & - \\ 1 & -1 & 0 \end{pmatrix}$$

(تصفيفة المشاهدات) = (تصفيفة المتوسط الكلي) + (تأثير تصفيفة المجتمعات (المعالجات)) + (تصفيفة البواقي)

لأن : $(x_{ki}) = (\bar{X}) + (\bar{X}_K - \bar{X}) + (x_{Ki} - \bar{X}_K)$

وبذلك نجد أن السؤال عن تساوي المتوسطات يتحول إلى تحديد فيما إذا كانت مساهمة تصفيفة تأثير المجتمعات (المعالجات) أكبر نسبياً من تصفيفة البواقي . علماً بأن التقديرات $\tilde{\tau}_k = (\bar{X}_k - \bar{X})$ للتأثيرات τ_k يجب أن تحقق دائماً الشرط (4–8) التالي :

$$\sum n_k\tilde{\tau}_k = 0$$

وضمن فرضية العدم H_0 ، يكون كل من $\tilde{\tau}_k$ تقديراً للعدد صفر، وهذا يعني أن تأثيرات المجتمعات تكون صغيرة .

أما إذا كانت تأثيرات المجتمعات (تأثيرات المعالجات) كبيرة . فإن ذلك سيؤدي إلى رفض الفرضية H_0 ، ولقياس مقدار مساهمة كل تصفيفة نكتب سطورها على شكل شعاع واحد، ثم نحسب مربع طول ذلك الشعاع، ويسمى هذا المقدار الجديد بمجموع المربعات (Sum of Squares) ونرمز له بالرمز SS .

فمثلاً نجد بالنسبة لتصفيفة القياسات المشاهدة (Observations) أن منقول شعاع عناصرها يكتب كمايلي:

$$Y' = [9\ 6\ 9\ 0\ 2\ 3\ 1\ 2]$$

وهو شعاع في الفضاء R^8 ، وإن مربع طوله يساوي $\|Y\|^2$ ويحسب كما يلي :

$$\|Y\|^2 = SS_{obs} = 9^2 + 6^2 + 9^2 + 0^2 + 2^2 + 3^2 + 1^2 + 2^2 = 216$$

وكذلك نجد أن مربع طول شعاع تصفيفة المتوسط الكلي يساوي:

$$SS_{means} = 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 + 4^2 = 128$$

وإن مربع طول شعاع تصفيفة تأثيرات المجتمعات أو (المعالجات treatments) يساوي :

$$SS_{tr} = 4^2 + 4^2 + 4^2 + (-3)^2 + (-3)^2 + (-2)^2 + (-2)^2 + (-2)^2 = 78$$

وإن مربع طول شعاع تصفيفة البواقي (residual) يساوي :

$$SS_{res} = 1^2 + (-2)^2 + 1^2 + (-1)^2 + 1^2 + 1^2 + (-1)^2 + 0^2 = 10$$

وبذلك نجد أن هذه المجاميع للمربعات ترتبط بنفس التركيب المعرف في (9−6) للمشاهدات . وهي تساوي :

$$SS_{obs} = SS_{means} + SS_{tr} + SS_{res}$$

حيث نلاحظ أن قيمها العددية تساوي :

$$216 = 128 + 78 + 10$$

أي أن تغيرات X الناتجة عن هذه المجتمعات تتوزع إلى ثلاث مركبات هي: تأثيرات المتوسطات والمعالجات والبواقي .

وإن تحليل التباين يقوم على مقارنة المقدار SS_{tr} (للمعالجات) مع المقدار SS_{res} (للبواقي) . فإذا كانت SS_{tr} أكبر بكثير من SS_{res} نرفض الفرضية H_0 ونقبل H_1 ونعتبر أن تغيرات X في هذه المجتمعات متباينة أو مختلفة .

والآن سنقوم باستخراج المعادلات الرياضية لهذه العلاقات وننطلق من العلاقة (9−6) التالية :

$$x_{ki} = \bar{X} + (\bar{X}_k - \bar{X}) + (x_{ki} - \bar{X}_k) \qquad (10-4)$$

ثم نطرح من الطرفين المتوسط الكلي $\bar{X}$ فنحصل على الانحرافات المصححة (أو المميزة) التالية :

$$(x_{ki} - \bar{X}) = 0 + (\bar{X}_k - \bar{X}) + (x_{ki} - \bar{X}_k)$$

وبذلك تختفي تصفيفة المتوسطات، ثم نقوم بتربيع الطرفين فنحصل على أن:

$$(x_{ki} - \bar{X})^2 = (\bar{X}_k - \bar{X})^2 + (x_{ki} - \bar{X}_k)^2 + 2(\bar{X}_k - \bar{X})(x_{ki} - \bar{X}_k) \qquad (11-4)$$

ثم نأخذ مجموع الحدود في الطرفين، المأخوذة على عناصر كل عينة k (أي $\sum_{i=1}^{n_k}$)، فنجد أنه ضمن كل عينة k يكون لدينا ما يلي :

$$\sum_{i=1}^{n_k}(x_{ki} - \bar{X})^2 = \sum_{i=1}^{n_k}(\bar{X}_k - \bar{X})^2 + \sum_{i=1}^{n_k}(x_{ki} - \bar{X}_k)^2 + 2(\bar{X}_k - \bar{X})\sum_{i=1}^{n_k}(x_{ki} - \bar{X}_k) \qquad (12-4)$$

وهنا نلاحظ أن المجموع الأول في الطرف الأيمن ليس له علاقة بدليل القياسات i , لذلك فهو يساوي:

$$\sum_{i=1}^{n_k}(\bar{X}_k - \bar{X})^2 = n_k(\bar{X}_k - \bar{X})^2$$

كما نلاحظ أن المجموع الأخير يساوي الصفر: $\sum_{i=1}^{n_k}(\mathrm{x}_{ki} - \bar{X}_k) = 0$

لأنه يمثل مجموع انحرافات قياسات X في العينة k عن متوسطها $\bar{X}_k$. وهذا يؤدي إلى انعدام الحد الأخير بكامله .

وبذلك تأخذ العلاقة (12−4) الشكل التالي :

$$\sum_{i=1}^{n_k}(x_{ki} - \bar{X})^2 = n_k(\bar{X}_k - \bar{X})^2 + \sum_{i=1}^{n_k}(x_{ki} - \bar{X}_k)^2 + 0 \qquad (13-4)$$

وذلك ضمن كل عينة k مسحوبة من ذلك المجتمع .

ثم نقوم بأخذ مجموع الحدود في الطرفين، المأخوذ على جميع المجتمعات $\left(\text{أي } \sum_{k=1}^{g}\right)$ فنجد أن:

$$\sum_{k=1}^{g}\sum_{i=1}^{n_k}(x_{ki}-\bar{X})^2=\sum_{k=1}^{g}n_k(\bar{X}_k-\bar{X})^2+\sum_{k=1}^{g}\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k)^2 \qquad (14-4)$$

وهذا يعني أن هذه الأطراف حسب المفاهيم السابقة تساوي ما يلي:

(مجموع مربعات البواقي داخل العينات) + (مجموع المربعات بين العينات المصحح) = (اجمالي مجموع المربعات المصحح بعد طرح $\bar{X}$)

ونرمز لهذه المجاميع بالرموز التالية :

$$SST=SSB+SSW \qquad (15-4)$$

وهنا نشير إلى أن درجة حرية الحد SSB تساوي $(g-1)$ لأن المتوسطات $\bar{X}_k$ ترتبط مع المتوسط العام $\bar{X}$ بالعلاقة $\left(\bar{X}=\frac{1}{n}\sum n_k\,\bar{X}_k\right)$، وهذا ما ينقص عدد درجات الحرية بمقدار واحد وتصبح $v_1=(g-1)$، ولإيجاد درجة حرية الحد الأخير SSW، نأخذ المجموع الداخلي منه $\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k)^2$ فنجد أن درجة حريته تساوي (n_k-1) لأن قياسات كل عينة x_{ki} مرتبطة بمتوسطها $\bar{X}_k$ وفق العلاقة $\left(\bar{X}_k=\frac{1}{n_k}\sum x_{ki}\right)$، ومن ذلك نجد أن درجة حرية المجموع المضاعف $\sum_{k=1}^{g}\left[\sum_{i=1}^{n_k}(x_{ki}-\bar{X}_k)^2\right]$ تساوي مجموع درجات الحرية لما بداخله، أي أنها تساوي :

$$v_2=\sum_{k=1}^{g}(n_k-1)=\sum_{k=1}^{g}n_k-g=n-g \qquad (16-4)$$

ولحساب درجة حرية الطرف الأيسر SST ، نأخذ مجموع درجات الحرية لحرية الطرف الأيمن، وبذلك تكون درجة حرية SST مساوية لما يلي : $v=v_1+v_2$ أي أن:

$$v=(g-1)+(n-g)=n-1 \qquad (17-4)$$

وأخيراً نضع نتائج هذه الحسابات في جدول منظم كالتالي :

جدول (4-2): الجدول النموذجي لـ ANOVA الأحادي One way .

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط مجموع المربعات	قيمة F
بين العينات (للمعالجات)	$SSB=\sum_{k=1}^{g}n_k(\bar{X}_k-\bar{X})^2$	$v_1=g-1$	$M\,SSB=\frac{SSB}{(g-1)}$	$F=\frac{M\,SSB}{M\,SSW}$
داخل العينات (الخطأ)	$SSW=\sum_{k}^{g}\sum_{i}(x_{ki}-\bar{X}_k)^2$	$v_2=\sum_{k=1}^{g}n_k-g$	$M\,SSW=\frac{SSW}{(\sum n_k-g)}$	
المجموع الكلي المصحح	$SST=\sum_{k}^{g}\sum_{i}^{n_k}(x_{ki}-\bar{X})^2$	$v=\sum_{k=1}^{g}n_k-1$		

ولاختبار صحة الفرضية H_0 نستخدم المؤشر F المعرف بالعلاقة التالية :

$$F=\frac{SSB/(g-1)}{SSW/(\sum n_k-g)}=\frac{M\,SSB}{M\,SSW} \qquad (18-4)$$

ثم نقارن هذه القيمة المحسوبة مع القيمة الحرجة $F_{v_1v_2}(\propto)$ لمتحول التوزيع F المقابلة لمستوى الدلالة $\propto$ ولدرجتي الحرية :

$$v_1 = g - 1 \qquad v_2 = \left(\sum_{k=1}^{g} n_k - g\right) = n - g$$

ونتخذ القرار كما يلي :

(19 − 4) إذا كانت $F \leq F_{v_1 v_2}(\propto)$ نقبل H_0 باحتمال ثقة قدره $(1 - \propto)$.

(20 − 4) أما إذا كانت $F > F_{v_1 v_2}(\propto)$ نرفض H_0 ونقبل H_1 بمستوى دلالة قدره $\propto$.

مثال (4−2): لنأخذ بيانات المثال (4−1) السابق والتي تتناول قياسات متحول واحد X من (3) عينات مسحوبة من (3) مجتمعات والتي كانت كما يلي:

$$\begin{matrix} P_1: (n_1 = 3) \\ P_2: (n_2 = 2): \\ P_3: (n_3 = 3) \end{matrix} \begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & - \\ 3 & 1 & 2 \end{pmatrix} : \begin{matrix} \bar{x}_1 = 8 \\ \bar{x}_2 = 1 \\ \bar{x}_3 = 2 \end{matrix} \quad , \quad \bar{X} = 4$$

ومنها نجد أن :

$$SSB = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 + n_3(\bar{X}_3 - \bar{X})^2$$
$$= 3(8-4)^2 + 2(1-4)^2 + 3(2-4)^2 = 78$$

$$SSW = \sum_{k=1}^{g}\sum_{i=1}^{n_g}(x_{ki} - \bar{X}_k)^2 = [(9-8)^2 + (6-8)^2 + (9-8)^2] +$$
$$+[(0-1)^2 + (2-1)^2] + [(3-2)^2 + (1-2)^2 + (2-2)^2] = 10$$

وبذلك نجد أن قيمة SST (المصححة) تساوي:

$$SST = 78 + 10 = 88$$

ثم ننظم جدولاً خاصاً بذلك كما يلي :

جدول (4−3): تحليل التباين ANOVA الأحادي One way .

مصدر التباين	مجموع مربعات الانحرافات	درجة الحرية	متوسط مجموع المربعات	قيمة F
بين العينات (للمعالجات)	$SSB = 78$	$g - 1 = 2$	$M\ SSB = \frac{78}{2} = 39$	$F = \frac{M\ SSB}{M\ SSW} = \frac{39}{2}$
داخل العينات (الخطأ)	$SSW = 10$	$\sum n_k - g = 5$	$M\ SSW = \frac{10}{5} = 2$	
التباين الكلي (المصحح)	$SST = 88$	$\sum n_k - 1 = 7$		

ومنه نجد أن قيمة F المحسوبة تساوي :

$$F = \frac{M\ SSB}{M\ SSW} = \frac{\frac{SSB}{g-1}}{\frac{SSW}{\sum n_k - g}} = \frac{\frac{78}{2}}{\frac{10}{5}} = \frac{39}{2} = 19.5$$

ومن جداول التوزيع F نجد أن القيمة الحرجة لـ $F_{v_1 v_2}(\propto)$ عندما تكون $\propto = 0,05$ والمقابلة لدرجتي الحرية $v_1 = 2$ و $v_2 = 5$ ، تساوي: $F_{2,5}(0,05) = 5.786$: وبمقارنة F المحسوبة مع $F_{2,5}(\propto)$ الحرجة نجد أن :

$$(F = 19,5) > \left(F_{2,5}(0,05) = 5.786\right)$$

لذلك نرفض فرضية العدم H_0 التي تقول أن: $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ بمستوى دلالة 0.05، ونقبل H_1 التي تقول بوجود فروقات بين متوسطات تلك المجتمعات . وهنا يجب أن نتابع البحث عن مصدر تلك الفروقات .

ملاحظة 1: مما سبق نستنتج أنه يتم رفض H_0 عندما تأخذ النسبة $\frac{M\,SSB}{M\,SSW}$ قيمة كبيرة (أكبر من $F(\propto)$)، أو عندما تكون قيمة المقدار $\left(1 + \frac{M\,SSB}{M\,SSW}\right)$ كبيرة أيضاً، وهذا يكافئ القول التالي: إننا نرفض H_0 عندما تكون قيمة مقلوب المقدار السابق صغيرة . أي أننا نرفض H_0 عندما تكون قيمة المقدار التالي :

$$\lambda = \frac{1}{1 + \frac{M\,SSB}{M\,SSW}} = \frac{M\,SSW}{M\,SSB + M\,SSW} \qquad (20\,a - 4)$$

صغيرة بقدر كاف لتحقيق مستوى الدلالة $\propto$.

ويستفاد من هذه العلاقة في إيجاد العلاقة , التي سوف نستخدمها في رفض H_0 في حالة عدة متحولات, كما سنرى لاحقاً .

ملاحظة 2: حول علاقة مؤشر تحليل التباين F باختبار (ستودينت) t : عندما يكون لدينا مجتمعان فقط $(g = 2)$, فإنه يمكننا كتابة الحد SSB كما يلي:

$$SSB = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 \qquad (20\,b - 4)$$

علماً بأن المتوسط الكلي $\bar{X}$ يحسب من العلاقة التالية :

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$$

وبتعويض $\bar{X}$ في $(20\,b - 4)$ نحصل على أن :

$$SSB = n_1\left(\bar{X}_1 - \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}\right)^2 + n_2\left(\bar{X}_2 - \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}\right)^2$$

$$SSB = \frac{n_1[(n_1 + n_2)\bar{X}_1 - (n_1\bar{X}_1 + n_2\bar{X}_2)]^2 + n_2[(n_1 + n_2)\bar{X}_2 - (n_1\bar{X}_1 + n_2\bar{X}_2)]^2}{(n_1 + n_2)^2}$$

$$SSB = \frac{n_1\; n_2(\bar{X}_1 - \bar{X}_2)^2}{n_1 + n_2} = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

وبذلك نجد أن المؤشر F (عندما $g = 2$) يأخذ الشكل التالي :

$$F = \frac{\frac{SSB}{g-1}}{\frac{SSE}{n-g}} = \frac{\frac{SSB}{1}}{\frac{SSE}{n_1 + n_2 - 2}} = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{SSE_1 + SSE_2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

وبما أن $SSE_1 = (n_1 - 1)S_1^2$ و $SSE_2 = (n_2 - 1)S_2^2$ نجد أن:

$$F_{1,n-2}(\propto) = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = t_{n-2}^2\left(\frac{\propto}{2}\right): \qquad (20\,C - 6)$$

وبمقارنة العبارة الأخيرة لـ F مع المؤشر t نجد: $F_{1,n-2}(\propto) = t_{n-2}^2\left(\frac{\propto}{2}\right)$ ، وهذا يعني أن مؤشر تحليل التباين F لمجتمعين يكافئ مربع اختبار (ستودينت) t، كما يمكن اعتبار الاختبار F تعميماً لاختبار (ستودينت) t .

2-4: تحليل التباين البسيط باتجاهين (وn مشاهدة) (ANOVA tow way) :

إن تحليل التباين البسيط باتجاهين يختلف جوهرياً عن تحليل التباين البسيط باتجاه واحد . وهو يستخدم لدراسة تغيرات متحول واحد (X) الناتجة عن تأثير عاملين نوعيين F_A و F_B (وليس عن مجتمعين كما في حالة الاتجاه الواحد)، وإن كل من هذين العاملين يأخذ عدة حالات، يدخلهما أو يتحكم بهما الباحث خلال مجربات التجارب، ثم يدرس تأثيرهما على المتحول التابع X. وهكذا يتم تنفيذ معظم الأبحاث العلمية.

فمثلاً: يمكن للباحث أن يدرس تغيرات أسعار طراز معين من الفساتين بتأثير عاملين: نوع القماش F_A وشكله الفني F_B . كما يمكنه أن يدرس تغيرات الجاذبية الأرضية بتأثير درجة الطول F_A ودرجة العرض F_B .

وفي هذه الحالة يتوجب على الباحث تحديد الحالات أو القيم التي يأخذها كل من F_A و F_B، وأن يرسم جدولاً خاصاً لتقاطعاتهما، ثم عليه أن يجري تجربة واحدة على الأقل مقابل كل حجرة لتقاطعهما . ثم عليه وضع نتائج تلك التجارب في جدول مناسب لحالات تقاطع F_A و F_B .

ولنفترض الآن أن F_A يأخذ g حالة منفصلة، وأن F_B يأخذ q حالة منفصلة . وان الباحث قد أجرى تجربة واحدة مقابل كل تقاطع لهما ووضع نتائجه في جدول كالتالي :

جدول (4-4): الحالات المتقاطعة لـ F_A و F_B

F_A / F_B	1	2	3	…	g
1	x_{11}	x_{21}	x_{31}	…	x_{g1}
2	x_{12}	x_{22}	x_{32}	…	x_{g1}
⋮	⋮	⋮	⋮	…	⋮
q	x_{1q}	x_{2q}	x_{3q}	…	x_{gq}

ولكن الباحث يمكن أن يقوم بتكرار تجاربه مقابل كل حجرة (k,ℓ) عدداً من المرات . وليكن n مرة، فعندها سيحصل في كل حجرة على n نتيجة . وهذه النتائج تمثل عينة من القياسات مأخوذة من مجتمع التجارب الذي يقابل كل حجرة (k,ℓ) . وبذلك يكون لدينا $(g * q)$ مجتمعاً إحصائياً سحبت منها $(g * q)$ عينة عشوائية بحجوم متساوية n قياساً لـ X في كل منها .

ولنفترض أن توقع X في كل حجرة (k,ℓ) منها يساوي $(\mu_{k\ell})$, وأن التوقع الكلي لـ X لها يساوي μ حيث أن:

$$\mu = \frac{\sum_{k=1}^{g}\sum_{\ell=}^{q}\mu_{k\ell}}{g*q} \qquad (21-4)$$

وعندها يمكننا أن نعبر عن التوقع $\mu_{k\ell}$ في الحجرة (k,ℓ) بعلاقة مركبة كما يلي :

$$\mu_{k\ell} = \mu + (\mu_{k\ell} - \mu) \qquad (22-4)$$

وحتى نظهر تأثير العاملين F_A وF_B وتأثيرهما المشترك $(F_A\ F_B)$ على نتائج تلك التجارب نكتب التوقع $\mu_{K\ell}$ على الشكل التالي (وذلك بإضافة وطرح التوقعات الهامشية μ_k وμ_ℓ من الطرف الأيمن) :

$$\mu_{k\ell} = \mu + (\mu_k - \mu) + (\mu_\ell - \mu) + (\mu_{k\ell} - \mu_k - \mu_\ell + \mu) \qquad (23-4)$$

وهنا نلاحظ أن الأقواس تعكس تأثير العامل F_A والعامل F_B وتأثيرهما معاً . واختصاراً للرموز نكتب ذلك كما يلي :

$$\mu_{k\ell} = \mu + \propto_k + \beta_\ell + \gamma_{k\ell} \qquad (24-4)$$

حيث أن: $\propto_k = (\mu_k - \mu)$ و $\beta_\ell = (\mu_\ell - \mu)$ و $\gamma_{k\ell} = (\mu_{k\ell} - \mu_k - \mu_\ell + \mu)$

وهنا يشترط على المقادير $\propto_k$ وβ_ℓ و$\gamma_{k\ell}$ أن تحقق الشروط التالية (يمكن البرهان على ذلك كما فعلنا في (4–8) مع ملاحظة أن حجوم العينات هنا متساوية وتساوي n) .

$$\sum_{k=1}^{g}\propto_k = 0 \qquad \sum_{\ell=1}^{q}\beta_\ell = 0$$
$$\sum_{k=1}^{g}\gamma_{k\ell} = 0 \qquad \sum_{\ell=1}^{q}\gamma_{k\ell} = 0 \qquad (25-4)$$

والآن نعود إلى العلاقة (4–24) ونكتبها كما يلي:

$$\mu_{k\ell} = E(x_{k\ell i}) = \mu + \propto_{\mathrm{k}} + \beta_\ell + \gamma_{k\ell} \qquad (26-4)$$

وفي العينات يمكننا كتابة قيمة أي قياس $x_{K\ell i}$ بخطأ $e_{K\ell i}$ كما يلي:

$$x_{k\ell i} = \bar{X} + \propto_k + \beta_\ell + \gamma_{k\ell} + e_{k\ell i} \qquad (27-4)$$

أو كما يلي :

$$x_{K\ell i} = \bar{X} + (\bar{X}_k - \bar{X}) + (\bar{X}_\ell - \bar{X}) + (\bar{X}_{K\ell} - \bar{X}_k - \bar{X}_\ell - \bar{X}) + (x_{K\ell i} - \bar{X}_{K\ell}) \qquad \mathbf{(28-4)}$$

(قيمة القياس $x_{K\ell i}$) = (المتوسط الكلي) + (تأثير العامل الأول F_A) + (تأثير العامل الثاني F_B) + (تأثير تداخل العاملين F_A و F_B معا) + (الخطأ العشوائي أو البواقي)

حيث أن $\bar{X}_k$ ترمز لمتوسطات العامل الأول F_A وأن $k : 1\ 2\ 3 \ldots g$

وأن $\bar{X}_\ell$ ترمز لمتوسطات العامل الثاني F_B وأن $\ell : 1\ 2\ 3 \ldots q$

وأن $\bar{X}_{k\ell}$ ترمز لمتوسطات القياسات في الحجرة (k,ℓ): $\bar{X}_{k\ell} = \frac{1}{n}\sum_{i=1}^{n} x_{k\ell i}$

وأن $e_{k\ell i}$ هي قيم البواقي أو الخطأ العشوائي، وهي عبارة عن متحولات عشوائية مستقلة ضمن كل حجرة وخاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$ الذي توقعه يساوي الصفر ولها تباين موحد يساوي σ^2 .

وعندها فإن الفرضيات البحثية تأخذ الشكل التالي: فرضية العدم: وهي تتألف مما يلي :

$$H_0 : \begin{cases} \propto_k = 0 & k : 1\ 2\ 3 \dots g \\ \beta_\ell = 0 & \ell : 1\ 2\ 3 \dots q \\ \gamma_{k\ell} = 0 & k,\ell : 1\ 2\ 3 \dots \end{cases} \qquad (29-4)$$

أما الفرضية البديلة فتكون كما يلي:

$$H_1 : \begin{cases} \propto_K \neq 0 & \text{من أجل } k \text{ واحدة على الأقل :} \\ \beta_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل :} \\ \gamma_{k\ell} \neq 0 & \text{من أجل زوج } (\ell , k) \text{ واحد على الأقل :} \end{cases} \qquad (30-4)$$

ولاستخراج مؤشرات الاختبار المناسبة نقوم بمعالجة العلاقة السابقة (4–28) كما فعلنا مع العلاقة السابقة (4–12) فنحصل على العلاقة التالية :

$$\sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{i=1}^{n}(x_{k\ell i}-\bar{X})^2 = qn\sum_{k=1}^{g}(\bar{X}_k-\bar{X})^2 + gn\sum_{\ell=1}^{q}(\bar{X}_\ell-\bar{X})^2 +$$

$$+n\sum_{k=1}^{g}\sum_{\ell=1}^{q}(\bar{X}_{k\ell}-\bar{X}_k-\bar{X}_\ell+\bar{X})^2 + \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{i=1}^{n}(x_{k\ell i}-\bar{X}_{k\ell})^2 \qquad (31-4)$$

والتي سنرمز لأطرافها اختصاراً كما يلي:

$$SST = SSA + SSB + SSAB + SSE \qquad (32-4)$$

أما بالنسبة لدرجات الحرية التي تقابل كل منها فهي تساوي:

$$(g*q*n-1) = (g-1)+(q-1)+(g-1)(q-1)+gq(n-1) \qquad (33-4)$$

ثم نعرف مؤشرات الاختبار المناسبة لكل من F_A و F_B وللتداخل بينهما كما يلي:

$$F_A = \frac{\frac{SSA}{g-1}}{\frac{SSE}{gq(n-1)}} = \frac{gq(n-1)SSA}{(g-1)SSE} \qquad \left(\text{للعامل } F_A\right) \qquad \mathbf{(34-4)}$$

وهو يخضع للتوزيع F بدرجتي حرية $v_1 = (g-1)$ و $v_2 = g*q(n-1)$.

$$F_B = \frac{\frac{SSB}{q-1}}{\frac{SSE}{gq(n-1)}} = \frac{gq(n-1)SSB}{(q-1)SSE} \qquad \left(\text{للعامل } F_B\right) \qquad \mathbf{(35-4)}$$

وهو يخضع للتوزيع F بدرجتي حرية $v_1 = (q-1)$ و $v_2 = g*q(n-1)$.

أما مؤشر اختبار التداخل بين F_A و F_B فيعرف كما يلي:

$$F_{AB} = \frac{\frac{SSAB}{(g-1)(q-1)}}{\frac{SSE}{gq(n-1)}} = \frac{gq(n-1)SSAB}{(g-1)(q-1)SSE} \qquad \mathbf{(36-4)}$$

وهو يخضع للتوزيع F بدرجتي حرية $v_1 = (g-1)(q-1)$ و $v_2 = g*q(n-1)$.

ثم نقوم بتنظيم جدول مناسب لتحليل التباين البسيط باتجاهين كما يلي:

جدول (4-5): جدول ANOVA باتجاهين tow way:

المؤشر F	درجة الحرية	مجموع المربعات	مصدر التباين
$F_A = \dfrac{gq(n-1)SSA}{(g-1)SSE}$	$g-1$	$SSA = (\quad\quad)$	العامل F_A
$F_B = \dfrac{gq(n-1)SSB}{(q-1)SSE}$	$q-1$	$SSB = (\quad\quad)$	العامل F_B
$F_{AB} = \dfrac{gq(n-1)SSAB}{(g-1)(q-1)SSE}$	$(g-1)(q-1)$	$SSAB = (\quad\quad)$	التداخل F_AF_B
———	$gq(n-1)$	$SSE = (\quad\quad)$	البواقي أو الخطأ العشوائي
———	$gq\,n-1$	$T=(\quad\quad)SS$	الإجمالي (المصحح)

وبعدها نقوم بمقارنة كل من F_A و F_B و F_{AB} بالقيم الحرجة المقابلة لها $F_{v_1,v_2}(\propto)$ بدرجتي الحرية v_1 و v_2 و نتخذ القرار حسب العامل المفروض كما يلي :

إذا كانت $F \le F_{v_1,v_2}(\propto)$ نقبل الفرضية H_0 حسب العامل المفروض

أما إذا كانت $F > F_{v_1,v_2}(\propto)$ نرفض الفرضية H_0 ونقبل H_1 $\quad(37-4)$

ثم نستخلص النتائج الممكنة من هذه الاختبارات كما سنرى من خلال المثال التالي :

مثال (4-3): لنفترض أننا نريد دراسة تغيرات أسعار طراز معين من الفساتين في السوق، ومعرفة درجة تأثر أسعارها X بنوعية القماش F_A أو بشكل وزخرفة الفستان F_B . وبعد الدراسة تبين لنا أن هذه الفساتين تُصنع من (3) أنواع من القماش هي $A(A_1, A_2, A_3)$، وأن شكلها يأخذ شكلين أساسين من الزخرفة هما (B_1, B_2) . ثم قمنا بتتبع أسعار هذه الفساتين حسب كل تقاطع لحالات النوع والشكل، ولذلك أخذنا (4) أسواق (محلات) تبيع هذه الفساتين وسجلنا الأسعار فيها حسب النوع والشكل, فكانت كما يلي (حسبنا متوسطات الاسعار في كل حجرة ووضعناها ضمن مستطيلات):

جدول (4-6) بيانات المثال (فرضية):

المتوسطات $X\ell$	الاجمالي		نوع ثالث A_3	نوع ثاني A_2	نوع أول A_1	أنواع القماش / أشكال الفساتين
447.5	5370		420 440 [450] 460 480	410 420 [425] 430 440	430 450 [467.5] 460 530	أحمر مزخرف B_1
392.5	4710		390 370 [390] 400 400	350 370 [380] 400 400	400 400 [407.5] 400 430	أبيض مزخرف B_2
	1008 0		3360	3220	3500	الاجمالي X_k
$\bar{X} = 420$			420	402.5	437.5	المتوسطات $\bar{X}_k$

والمطلوب دراسة تأثير العاملين A وB على السعر X . وإجراء الاختبارات اللازمة بمستوى دلالة =∝ 0.05) ، علماً بأن الأرقام ضمن المستطيلات في كل حجرة هي متوسطات القياسات فيها $\bar{X}_{k\ell}$ وإن المتوسط الكلي لها $\bar{X} = 420$.

الحل: نقوم أولاً بحساب المجاميع التي في العلاقة (6-31) فنجد أن:

$$SST = \sum_{k=1}^{3}\sum_{\ell=1}^{2}\sum_{i=1}^{4}(x_{k\ell i} - 420)^2 = 34600 \qquad (31-4)$$

وذلك لأن:

$$\begin{aligned} SST = {} & (100 + 900 + 1600 + 12100) + (100 + 0 - 100 + 400) + (0 + 400 + 1600 + 3600) \\ & + (400 + 400 + 400 + 100) + (4900 + 2500 + 400 + 400) \\ & + (900 + 2500 + 400 + 400) = 34600 \end{aligned}$$

ثم نقوم بحساب SSA من العلاقة :

$$SSA = qn\sum_{k=1}^{3}(\bar{X}_k - \bar{X})^2 =$$

$$SSA = 2 * 4[306.5 + 306.5 + 0] = 4900$$

ثم نقوم بحساب SSB من العلاقة :

$$SSB = gn\sum_{\ell=1}^{2}(\bar{X}_\ell - \bar{X})^2 =$$

$$SSB = 3 * 4[756.25 + 756.25] = 18150$$

ثم نقوم بحساب حد البواقي أو الخطأ العشوائي في جميع الحجر من العلاقة :

$$SSE = \sum_{k=1}^{3}\sum_{\ell=1}^{2}\sum_{i=1}^{4}(x_{k\ell i} - \bar{X}_{k\ell})^2 =$$

$$SSE = 5675 + 500 + 2000 + 675 + 1800 + 600 = 11250$$

وأخيراً نقوم بحساب حد التداخل SSAB من العلاقة :

$$SSAB = SST - SSA - SSB - SSE$$

$$SSAB = 34600 - 4900 - 18150 - 11250 = 300$$

ثم نقوم بتنظيم جدول التحليل التالي :

جدول (4-7) جدول تحليل ANOVA باتجاهين :

مصدر التباين	مجموع المربعات	درجة الحرية	متوسطات المربعات	قيم F
نوعية القماش A	$SSA = 4900$	$g - 1 = 2$	$MSSA = 2450$	$FA = 3.92$
شكل الفستان B	$SSB = 18150$	$q - 1 = 1$	$MSSB = 18150$	$F_B = 29.04$
تداخل A وB	$SSAB = 300$	$(g-1)(q-1) = 2$	$MSSAB = 150$	$F_{AB} = 0.24$
الخطأ العشوائي	$SSE = 11250$	$gq(n-1) = 18$	$MSSE = 625$	ـــــــ
الاجمالي	$SST = 34600$	$gqn - 1 = 19$	ـــــــ	ـــــــ

ولاختبار تأثير هذين العاملين وتأثير تداخلهما، علينا أن نقوم بإيجاد القيمة الحرجة لـ F المقابلة لكل حالة فنجد أن:

$$F_{1,18}(0.05) = 4.41 \quad \text{و} \quad F_{2,18}(0.05) = F_{2,18}(0.05) = 3.55$$

ثم نقوم بمقارنة قيم F_A مع $F_{2,18}(\propto)$ نجد أن $(F_A = 3.92) > 3.55$: , لذلك نرفض فرضية العدم H_0، التي تقول أنه لا يوجد تأثير لنوعية القماش على سعره , ونقبل H_1 التي تقول أن أسعار الفساتين تتأثر بنوعية القماش (بنسبة $\left(\frac{4900}{34600}\right) * 100\%$) .

ثم نقوم بمقارنة F_B مع $F_{1,18}(\propto)$ نجد أيضاً أن $(F_B = 29.04) > 4.41$,: , لذلك نرفض فرضية العدم التي تقول أنه لا يوجد تأثير لشكل الفستان على سعره , ونقبل H_1 التي تقول أن أسعار الفساتين تتأثر كثيراً بشكل القماش (بنسبة $\left(\frac{18150}{34600}\right) * 100\%$) .

ولاختبار تأثير التفاعل الداخلي للعاملين A وB نقارن قيمة F_{AB} مع القيمة الحرجة $F_{2\ 18}(\propto)$ فنجد أن : $(F_{AB} = 0.24) < 3.55$, لذلك نقبل H_0 ونعتبر أن أسعار الفساتين لا تتأثر بتداخل العاملين المدروسين وهما النوعية والشكل .

ملاحظة: يفضل في التطبيقات العملية أن نبدأ بإجراء اختبار التداخل . فإذا كان تأثير ذلك التداخل معنوياً (حالة رفض H_0) . فإن ذلك يعني أن تداخل العاملين F_A وF_B يؤثر كثيراً على تغيرات المتحول التابع X . وهذا يجعل عملية اختبار تأثير كل من F_A وF_B على حدة على X غير واضحة وصعبة التفسير، وينصح بعدم متابعة دراسة تأثيراتهما المنفردة.

أما إذا كان تأثير التداخل مهملاً فإنه يمكننا متابعة التحليل وإجراء الاختبارين حول تأثير F_A وF_B واستخلاص النتائج الممكنة .

4-3: تحليل التباين البسيط بثلاث اتجاهات (n مشاهدات) :

إن تحليل التباين في هذه الحالة هو تعميم للحالة السابقة (4-2) . لذلك نفترض أننا نريد دراسة تغيرات أحد المتحولات X (المتحول التابع) الناتجة عن تأثير (3) عوامل نوعية: F_A وF_B وF_C . وإن لكل من هذه العوامل عدة حالات يرمز لها كما يلي:

$$\begin{aligned} F_A &: A_1, A_2, \dots, A_K, \dots, A_g \\ F_B &: B_1, B_2, \dots, B_\ell, \dots, B_q \\ F_C &: C_1, C_2, \dots, C_m, \dots, C_r \end{aligned} \qquad (38-4)$$

ولذلك ننشأ مكعب التقاطعات الممكنة لهذه الحالات، فنحصل على $(g * q * r)$ حجرة، ثم نأخذ من كل حجرة منه n قياساً للمتحول المدروس X . فنحصل على $(g * q * r)$ عينة مسحوبة من مجتمعات تلك الحجر وعلى $(g * q * r * n)$ قياساً .

وإذا رمزنا لتلك القياسات بالرموز $x_{k\ell mi}$ فإنه يمكننا كتابة النموذج الرياضي الموافق لهذا التحليل باستخدام نفس المفاهيم والرموز السابقة كما يلي :

$$x_{k\ell mi} = \bar{X} + \propto_k + \beta_\ell + \gamma_m + (\propto\beta)_{k\ell} + (\propto\gamma)_{km} + (\beta\gamma)_{\ell m} + (\propto\beta\gamma)_{k\ell m} + (e_{k\ell mi}) \quad \mathbf{(39-4)}$$

ويشترط على هذه الرموز أن تحقق شروط مشابهة للشروط السابقة المفروضة على تحليل التباين باتجاهين والمذكورة في (4-25) .

أما فرضية العدم H_0 (وهي عدم وجود تأثير لهذه العوامل على X) فتكتب كما يلي:

$$H_0:\begin{cases}\propto_1=\propto_2=\cdots=\propto_g=0\\ \beta_1=\beta_2=\cdots=\beta_q=0\\ \gamma_1=\gamma_2=\cdots=\gamma_r=0\end{cases} \qquad (40-4)$$

أما الفرضية البديلة H_1 فتكتب كما يلي:

$$H_1:\begin{cases}\propto_k\neq 0 & \text{من أجل } k \text{ واحدة على الأقل}\\ \beta_\ell\neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل}\\ \gamma_m\neq 0 & \text{من أجل } m \text{ واحدة على الأقل}\end{cases} \qquad (41-4)$$

ثم نقوم بمعالجة العلاقة (4-39) كما فعلنا في الفقرات السابقة فنحصل مجاميع المربعات المختلفة ونكتبها كما يلي:

$$SST=SSA+SSB+SSC+SS(AB)+SS(AC)+SS(BC)+SS(ABC)+SSE \quad (42-4)$$

وإن درجات الحرية المقابلة لهذه المجاميع هي كما يلي:

$$(gqrn)-1=(g-1)+(q-1)+(r-1)+(g-1)(q-1)+(g-1)(r-1)$$
$$+(q-1)(r-1)+(g-1)(q-1)(r-1)+gqr(n-1) \qquad (43-4)$$

وسنقوم بتعريف هذه المجاميع وشرح كيفية حسابها لاحقاً من خلال المثال (4-4) :

أما بالنسبة لإجراء الاختبارات اللازمة نبدأ بإجراء اختبار تأثير التداخل الثلاثي SS(ABC)، ونحسب قيمة مؤشر الاختبار F_{ABC} المقابل له من العلاقة :

$$F_{ABC}=\frac{\dfrac{SS(ABC)}{(g-1)(q-1)(r-1)}}{\dfrac{SSE}{gqr(n-1)}}=\frac{\dfrac{SS(ABC)}{v_1}}{\dfrac{SSE}{v_2}} \qquad (44-4)$$

ثم نقارن هذه القيمة المحسوبة مع القيمة الحرجة لمتحول $F_{v_1,v_2}(\propto)$ المقابل لمستوى الدلالة $\propto$ ولدرجتي الحرية v_1 و v_2، ونتخذ القرار كما يلي:

إذا كانت $F\leq F_{v_1,v_2}(\propto)$ نقبل فرضية العدم H_0

وإذا كانت $F>F_{v_1,v_2}(\propto)$ نرفض H_0 ونقبل H_1 $\qquad (45-4)$

وإذا كانت النتيجة رفض H_0 وقبول H_1، فإن ذلك يعني أن تداخل العوامل الثلاثة يؤثر معنوياً على تغيرات المتحول المدروس X . لذلك فإن عملية البحث عن تأثير كل منها بمفرده تصبح غير واضحة وغير ممكنة التفسير .

أما إذا كانت نتيجة الاختبار السابق قبول H_0، فإن ذلك يعني أن تأثير تداخل العوامل الثلاثة غير معنوي ويمكن إهماله . وبعدها ننتقل إلى اختبارات تأثيرات التداخلات الثنائية . وإذا كانت غير معنوية، نقوم باختبار تأثيرات العوامل المنفردة وذلك بتطبيق نفس الإجراءات السابقة .

مثال (4-4): لدراسة مقاومة ألواح السيراميك للصدمات ثم تحديد (3) عوامل مؤثرة عليها وهي :

- نوع المادة الأولية وتأخذ حالتين فقط (A_1, A_2) .
- مساحة اللوح ويأخذ حالتين أيضاً (B_1, B_2) .
- سماكة اللوح ويأخذ (3) حالات هي : (C_1, C_2, C_3) .

ثم تم اختبار (5) ألواح من كل حجرة لتقاطعاتها وأجريت عليها تجارب لقياس المقاومة (بالكيلوغرام) فكانت نتائج تلك القياسات كما في الجدول التالي :

جدول (4-7): بيانات المثال (فرضية)

نوع المادة		نوع المادة A_1						نوع المادة A_2					
المساحة		المساحة B_1			المساحة B_2			المساحة B_1			المساحة B_2		
السماكة		C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
عناصر العينة i	$i=1$	67	72	78	67	79	78	54	52	63	54	57	60
	2	66	67	81	71	80	78	51	56	54	56	58	68
	3	62	75	67	72	81	77	47	52	65	58	61	61
	4	71	70	75	70	80	83	51	52	62	51	59	61
	5	69	71	75	81	85	79	59	53	60	57	55	67

والمطلوب دراسة تأثير هذه العوامل على مقاومة ألواح السيراميك وإجراء الاختبارات اللازمة عليها بمستوى دلالة $\propto = 0.05$

الحل: لإجراء هذه الاختبارات نحتاج إلى كثير من الحسابات المعقدة لإيجاد قيم حدود العلاقة (4-42) وسنستخدم لذلك علاقات رياضية مبسطة (بدون برهانها) . وحتى نستطيع تطبيق تلك العلاقات أعددنا الجدولين (4-8) و (4-9) واستخدمنا بياناتهما في عملية الحساب كما يلي:

نقوم بحساب SST من العلاقة التالية :

$$SST = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{m=1}^{r}\sum_{i=1}^{n}(x_{k\ell mi} - \bar{X})^2 = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{m=1}^{r}\sum_{i=1}^{n} x_{k\ell mi}^2 - \frac{X^2}{g*q*r*n} \qquad (45-4)$$

ومن الجدولين (4-8) و(4-9) نجد أن :

$$SST = 265174 - \frac{(3942)^2}{2*2*3*5} = 265174 - 258989.4 = 6184.6$$

ومن السطر الأخير في الجدول (4-8) نجد مباشرة أن :

$$SSE = \sum_{k=1}^{g}\sum_{\ell=1}^{q}\sum_{m=1}^{r} SSE_{k\ell mi} = 614$$

ثم نقوم بحساب SSA من بيانات الجدول (4-9) وباستخدام العلاقة التالية:

$$SSA = \frac{1}{qrn}\sum_{k=1}^{2} X_k^2 - \frac{X^2}{gqrn} = \frac{1}{30}[(2228)^2 + (2044)^2] - 258989.4 = 4403.3 \qquad (46-4)$$

جدول (4-8) نتائج الحسابات:

النوع K	نوع المادة A_1						نوع المادة A_2						
المساحة ℓ	المساحة B_1			المساحة B_2			المساحة B_1			المساحة B_2			
m السماكة	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	المجموع
i=1	67	72	78	67	79	78	54	52	63	54	57	60	
2	66	67	81	71	80	78	51	56	54	56	58	68	
3	62	75	67	72	81	77	47	52	65	58	61	61	
4	71	70	75	70	80	83	51	52	62	51	59	61	
5	69	71	75	81	85	79	59	53	60	57	55	67	
$X_{k\ell m}$ المجموع	335	355	377	361	405	395	362	265	304	276	290	311	3942
$\frac{X_{k\ell m}^2}{n}$	22445	25205	28425.6	26064.2	32803	31205	13728.8	14045	18483.2	15235.2	16820	200988	
$\sum x_{k\ell m}^2$	22491	25239	26535	26173	32827	31227	13808	14057	18554	15266	16840	20155	$T = 265174$
$SSE_{k\ell m}$	46	36	109.2	1196	22	22	79.2	12	70.8	30.8	20	57.2	$SSE = 614$

الجدول (4-9) مجاميع المجاميع حسب الحالات السابقة:

المجاميع	$m=1$	$m=2$	$m=3$	المجموع $X_{k\ell 1}$
X_{111}	335	355	377	= 1067
X_{121}	361	405	395	= 1161
X_{211}	362	265	304	= 831
X_{221}	276	290	317	= 883
$X_{k1} = \begin{cases} k=1 \\ k=2 \end{cases}$	696 538	760 555	772 621	= 2228 = 2044
$X_{\ell=2} = \begin{cases} \ell=1 \\ \ell=2 \end{cases}$	597 637	620 695	681 712	= 1898 = 2044
X_m	1234	1315	1393	$X = 3942$

ثم نقوم بحساب SSB بناءً على الجدول (4-9) وباستخدام العلاقة التالية:

$$SSB = \frac{1}{grn}\sum_{\ell=1}^{2} X_\ell^2 - \frac{X^2}{gqrn} = \frac{1}{30}[(1898)^2 + (2044)^2] - 258989.4 = 355.3 \qquad (47-4)$$

ثم نقوم بحساب SSC بناءً على الجدول (4-9) من العلاقة التالية :

$$SSC = \frac{1}{gqn}\sum X_m^2 - \frac{X^2}{gqrn} = \frac{1}{20}[(1234)^2 + (1315)^2 - (1393)^2] - 258989.4 = 632.1 \qquad \mathbf{(48-4)}$$

ثم نقوم بحساب SS(AB) بناءً على العمود الأخير الجدول (4-9) من العلاقة التالية:

$$SS(AB) = \frac{1}{rn}\sum_{k=1}^{2}\sum_{\ell=1}^{2} X_{k\ell}^2 - \frac{X^2}{gqrn} - SSA - SSB \qquad (49-4)$$

$$= \frac{1}{15}[(1067)^2 + (1161)^2 + (831)^2 + (883)^2] - 258989.4 - 4403.3 - 355.3 = 29.3$$

ثم نقوم بحساب SS(AC) بناءً على الجدول (4-9) من العلاقة التالية :

$$SS(AC) = \frac{1}{qn}\sum_{k=1}^{2}\sum_{m=1}^{2} X_{km}^2 - \frac{X^2}{gqrn} - SSA - SSC \qquad (50-4)$$

$$= \frac{1}{10}[(696)^2 + (760)^2 + (772)^2 + (538)^2 + (555)^2 + (621)^2] - 258989.4 - 4403.3 - 632.1 = 68.2$$

ثم نقوم بحساب SS(BC) بناءً على الجدول (4-9) من العلاقة التالية:

$$SS(BC) = \frac{1}{gn}\sum_{\ell=1}^{2}\sum_{m=1}^{3} X_{\ell m}^2 - \frac{X^2}{gqrn} - SSB - SSC \qquad (51-6)$$

$$= \frac{1}{10}[(597)^2 + (620)^2 + (681)^2 + (637)^2 + (695)^2 + (712)^2] - 258989.4 - 355.3 - 632.1 = 54$$

وأخيراً نحسب حد التداخل الثلاثي SS(ABC) من العلاقة (4-42) فنجد أن :

$$SS(ABC) = 6184.6 - 614 - 4403.3 - 355.3 - 632.1 - 29.3 - 68.2 - 54 = 10.4$$

ثم نقوم بوضع نتائج هذه الحسابات في جدول تحليل التباين ذي الاتجاهات الثلاثة المرفقة بدرجات الحرية كمايلي:

جدول (4-10): جدول تحليل التباين بثلاث اتجاهات :

مصدر التباين	رمز المصدر	درجة الحرية	مجموع المربعات	متوسط المربعات	$\tilde{F}$	$F(\propto)$
F_A نوع المادة	SSA	$g-1=1$	4403.3	4403.3	344	4.04
F_B المساحة	SSB	$q-1=1$	355.3	355.3	27.8	4.04
F_C السماكة	SSC	$r-1=2$	632.1	316.0	24.7	3.19
أثر المادة والمساحة	$SS(AB)$	$(g-1)(q-1)=1$	29.3	29.3	2.29	4.04
أثر المادة والسماكة	$SS(AC)$	$(g-1)(r-1)=2$	86.2	43.1	3.37	3.19
أثر المساحة والسماكة	$SS(BC)$	$(q-1)(r-1)=2$	54.0	27.0	2.11	3.19
أثر العوامل الثلاثة المادة والمساحة والسماكة	$SS(ABC)$	$(g-1)(q-1)(r-1)=2$	10.4	5.2	0.41	3.19
الخطأ العشوائي (البواقي)	SSE	$gqr(n-1)=48$	614.0	12.79	–	–
الإجمالي	SST	$(gqrn)-1=59$	6184.6	–	–	–

وبعد حساب قيم F لجميع هذه المجاميع نقارنها مع قيمة $F(\propto)$ الحرجة المقابلة لها, والمبينة في العمود الأخير من الجدول (4-10) فنلاحظ ما يلي:

1- إن تأثير التداخل الثلاثي غير معنوي لأن $F_{ABC} = 0.41 < 3.19$.

2- إن تأثير التداخل الثنائي SS(AC) معنوي وهذا يؤثر على تفسير النتائج .

3-إن تأثير التداخلين الثنائيين(SS(AB و (SS(BC غير معنويين .

4-إن تأثير العوامل الثلاثة F_A و F_B و F_C معنوية وأن العامل F_A هو الأكثر تأثيراً على تغيرات المقاومة .

4-4: تحليل المربع اللاتيني (بمشاهدة واحدة):

لقد لاحظنا في الفقرة السابقة أن تحليل التباين بثلاث اتجاهات (وn مشاهدة) يحتاج إلى حسابات معقدة وإلى عدد كبير من المشاهدات . ولكن يمكننا في بعض الأبحاث تنظيم وترتيب العوامل الداخلة في تحليل التباين الثلاثي وعرضها على شكل مربع يسمى المربع اللاتيني . وهو يستخدم كثيراً في الأبحاث العلمية كالأبحاث الزراعية والطبية والاقتصادية وغيرها . وهو يتألف من عدة عناصر هي :

- العامل الأول F_A ويأخذ g حالة منفصلة .
- العامل الثاني F_B ويأخذ أيضاً g حالة منفصلة .
- المربع اللاتيني الذي يتم تشكيله من تقاطع حالات العاملين F_A و F_B . وهو يتألف من g^2 خلية .
- جملة من طرائق المعالجة وعددها يساوي g طريقة أيضاً وذلك لتطبيقها على خلايا المربع اللاتيني، بحيث يتم تطبيق كل طريقة مرة واحدة في كل سطر، ومرة واحدة في كل عمود .
- متحول تابع X يقيس نتائج المعالجات السابقة في جميع الخلايا بواحدة قياس موحدة .

ولتمثيل أحد أشكال المربع اللاتيني ونفترض أن عدد حالات F_A و F_B يساوي $g = 3$ ، وأنه لدينا (3) طرائق هي $A\ B\ C$ لتطبيقها على تقاطعات تلك الحالات، مرة واحدة في كل سطر ومرة واحدة في كل عمود، وعندها يمكن أن يأخذ المربع اللاتيني الشكل التالي :

	F_{A1}	F_{A2}	F_{A3}
F_{B1}	A	B	C
F_{B2}	B	C	A
F_{B3}	C	A	B

الشكل (4- 2)

وهنا نلاحظ أن الشكل السابق (4- 2) للمربع اللاتيني هو أحد الأشكال الممكنة، وتم الحصول عليه بتطبيق تسلسل معين للطرائق المستخدمة هو $(A\ B\ C)$ على خلايا السطر الأول، ثم القيام بسحب عناصره إلى اليسار خطوة واحدة وتطبيقها على خلايا السطر الثاني (مع نقل A إلى الحجرة الأخيرة) . ثم سحب عناصره إلى اليسار خطوة واحدة وتطبيقها على خلايا السطر الثالث (مع نقل B إلى الحجرة الأخيرة) . وبذلك نحصل على ما يسمى الشكل القياسي للمربع اللاتيني، وهو المربع الذي يتم تشكيله بتدوير عناصر السطر الأول لتوزيعها على السطر الثاني ثم تدوير عناصر الثاني لتوزيعها على الثالث، وهكذا دواليك . وبذلك تكون عناصر الأسطر متناظرة مع عناصر الأعمدة. ولكن الشكل القياسي للمربع اللاتيني ليس وحيداً ,فهناك عدد من الأشكال التي تعطينا توزيعات مختلفة , حيث نجد أنه يمكننا ترتيب عناصر السطر الأول عشوائياً بـ !3 طريقة,

ثم ترتيب العناصر المتبقية في العمود الأول عشوائياً(بعد السطر الأول) بـ $2!$ طريقة وتبقى الخلية الأخيرة والتي ترتب بطريقة واحدة فقط . وبذلك يكون عدد الأشكال الممكنة (حسب الأسطر أو الأعمدة) في الحالة التي يكون فيها $n = 3$ مساوياً لـ: $m = 3! * 2! * 1 = 12$, وهذه الأشكال الممكنة تعطينا المربعات اللاتينية التالية :

$$\begin{bmatrix} A & B & C \\ B & C & A \\ C & A & B \end{bmatrix} \begin{bmatrix} A & C & B \\ C & B & A \\ B & A & C \end{bmatrix} \begin{bmatrix} B & A & C \\ A & C & B \\ C & B & A \end{bmatrix} \begin{bmatrix} B & C & A \\ C & A & B \\ A & B & C \end{bmatrix} \begin{bmatrix} C & A & B \\ A & B & C \\ B & C & A \end{bmatrix} \begin{bmatrix} C & B & A \\ B & A & C \\ A & C & B \end{bmatrix}$$

$$\begin{bmatrix} A & B & C \\ C & A & B \\ B & C & A \end{bmatrix} \begin{bmatrix} A & C & B \\ B & A & C \\ C & B & A \end{bmatrix} \begin{bmatrix} B & A & C \\ C & B & A \\ A & C & B \end{bmatrix} \begin{bmatrix} B & C & A \\ A & B & C \\ C & A & B \end{bmatrix} \begin{bmatrix} C & A & B \\ B & C & A \\ A & B & C \end{bmatrix} \begin{bmatrix} C & B & A \\ A & C & B \\ B & A & C \end{bmatrix}$$

ولكننا في الأبحاث العلمية لا نحتاج إلى جميع هذه الأشكال. بل نقوم باختيار أحدها عشوائياً . وبذلك تكون نتيجة التجارب (طرائق المعالجة) في كل خلية هي عبارة عن متحول عشوائي معرف عليها .

وهكذا نجد أن المربع اللاتيني هو تحليل ثلاثي الاتجاه (بمشاهدة واحدة) وإن العوامل المعتمدة فيه هي :

- عامل الأسطر ونرمز له بـ F_A وعدد حالاته g حالة .
- عامل الأعمدة ونرمز له بـ F_B وعدد حالاته g حالة أيضاً .
- طرائق المعالجة وعددها g حالة أيضاً. وهي تطبق على خلايا المربع اللاتيني، بحيث يتم تطبيق كل طريقة مرة واحدة في كل سطر ومرة واحدة في كل عمود, وحسب أحد الأشكال الممكنة للمربع.

ونتيجة لتطبيق هذه الطرائق على الخلايا نحصل على g^2 قياساً للمتحول التابع X، وسنرمز لقيمة المتحول التابع X في الخلية (k , ℓ) بالرمز $x_{k\ell}$ ونضعها إلى جانب رمز الطريقة المطبقة في تلك الخلية، ثم نحسب مجاميع ومتوسطات الأسطر والأعمدة ونضعها في الهوامش.

وإذا أخذنا الحالة التي يكون لدينا فيها: $(g = 4)$ طرائق ويكون لكل من العاملين F_A و F_B أربعة حالات، فإننا سنحصل على الجدول التالي:

جدول (4-11): المربع اللاتيني 4×4

الأعمدة F_A / الأسطر F_B	1	2	3	4	المجموع X_k	المتوسط $\bar{X}_k$
1	A x_{11}	B x_{12}	C x_{13}	D x_{14}	X_1	$\bar{X}_1$
2	B x_{21}	C x_{22}	D x_{23}	A x_{24}	X_2	$\bar{X}_2$
3	C x_{31}	D x_{32}	A x_{33}	B x_{34}	X_3	$\bar{X}_3$
4	D x_{41}	A x_{42}	B x_{43}	C x_{44}	X_4	$\bar{X}_4$
المجموع X_ℓ	X'_1	X'_2	X'_3	X'_4	X	$\bar{X}$
المتوسط $\bar{X}_\ell$	$\bar{X}'_1$	$\bar{X}'_2$	$\bar{X}'_3$	$\bar{X}'_4$	$\bar{X}$	$\bar{X}$

حيث أن هذه المجاميع والمتوسطات تحسب من العلاقات التالية :

$$\bar{X}_k = \frac{1}{n}\sum_{\ell=1}^{g} x_{k\ell} = \frac{X_k}{g} \text{ : } k \text{ المتوسط في السطر} \qquad X_k = \sum_{\ell=1}^{g} x_{k\ell} \text{ } k \text{ اجمالي السطر}$$

$$(52-4) \qquad \bar{X}_\ell = \frac{1}{n}\sum_{k=1}^{g} x_{k\ell} = \frac{X_\ell}{g} \text{ : } \ell \text{ المتوسط في العمود} \qquad X_\ell = \sum_{k=1}^{g} x_{k\ell} \text{ } \ell \text{ اجمالي العمود}$$

$$\text{المتوسط في الجدول ككل : } \bar{X} = \frac{X}{g*g} = \frac{\sum^g \sum^g x_{k\ell}}{g*g} \text{ ,} \qquad X = \sum_{k=1}^{g} X_k = \sum_{\ell=1}^{g} X_\ell = \sum_{k=1}^{g}\sum_{\ell=1}^{g} x_{k\ell} \text{ الاجمالي الكلي}$$

وهناك متوسطات من نوع آخر هي متوسطات قياسات X حسب طرائق المعالجة $(A\ B\ C\ D)$ ويتم حساب هذه المتوسطات لكل طريقة على حدة، وذلك بتتبع قيم X حسب كل طريقة ضمن المربع اللاتيني, ونرمز لهذه المتوسطات بالرموز $\bar{X}_t$، حيث t هو دليل الطريقة t، فمثلاً نجد أن متوسط قيم X المقابلة للطريقة A يحسب من الخلايا التي تطبق فيها الطريقة A وهو يساوي (حسب الجدول السابق (4-11)) مايلي:

$$\bar{X}_A = \frac{1}{4}[x_{11} + x_{24} + x_{33} + x_{42}] = \frac{X_A}{g}$$

وكذلك نجد أن متوسطات الطرائق الأخرى تساوي :

$$\bar{X}_B = \frac{1}{4}[x_{12} + x_{21} + x_{34} + x_{43}] = \frac{X_B}{g}$$

$$\bar{X}_C = \frac{1}{4}[x_{13} + x_{22} + x_{31} + x_{44}] = \frac{X_C}{g} \qquad (53-4)$$

$$\bar{X}_D = \frac{1}{4}[x_{14} + x_{23} + x_{32} + x_{41}] = \frac{X_D}{g}$$

ويمكن تسهيل حساب هذه المتوسطات بتنظيم جدول خاص بذلك، يتضمن تبويب قياسات X حسب طرائق المعالجة، وحسب الأسطر (أو حسب الأعمدة) كما يلي:

جدول (4-12): حساب المتوسطات حسب طرائق المعالجة

الطريقة t / الأسطر	A	B	C	D	X_i
1	x_{11}	x_{12}	x_{13}	x_{14}	X_1
2	x_{24}	x_{21}	x_{22}	x_{23}	X_2
3	x_{33}	x_{34}	x_{31}	x_{32}	X_3
4	x_{42}	x_{43}	x_{33}	x_{41}	X_4
المجموع	X_A	X_B	X_C	X_D	X
المتوسط $\bar{x}_t$	$\bar{X}_A$	$\bar{X}_B$	$\bar{X}_C$	$\bar{X}_D$	

وهنا يشترط على هذه القياسات في كل عمود (أو سطر) ان تكون مستقلة عن بعضها البعض ، وخاضعة للتوزيع الطبيعي ولها تباين موحد مساوٍ لـ σ^2 .

ولنفترض الآن أن التوقع الرياضي لقيم X في الخلية (k , ℓ) من المربع يساوي $\mu_{k\ell}$، وأن التوقع العام لقيم X في كامل المربع بالرمز μ . وبما أن القياسات مستقلة فإن التفاعلات الثنائية للعوامل تكون معدومة، كما أن التفاعل الثلاثي للعوامل الثلاثة معاً يكون معدوماً أيضاً .

وبناء على ذلك يمكننا أن نصيغ نموذج تحليل المربع اللاتيني لهذه العوامل الثلاثة المستقلة كما يلي:

$$\mu_{k\ell} = \mu + \propto_k + \beta_\ell + \gamma_t \qquad (54-4)$$

حيث أن: $\propto_k$ هو تأثير السطر k وβ_ℓ تأثير العمودℓ، وγ_t تأثير الطريقة t، وبما أن المتوسط العام في العينات هو تقدير لـ μ في المجتمع، فإنه يمكننا التعبير عن قيمة X في كل خلية كما يلي:

$$x_{k\ell} = \bar{X} + \propto_k + \beta_\ell + \gamma_t + e_{k\ell} \qquad (54\,a-4)$$

حيث أن: $e_{k\ell}$ هو حد الخطأ العشوائي ويخضع للتوزيع الطبيعي $N(0\,,\sigma^2)$، وهكذا نجد أن مجموع مربعات الانحرافات عن μ يساوي :

$$\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\mu)^2 \qquad (55-4)$$

ولإظهار تأثير العوامل الثلاثة نستبدل μ بتقديره $\bar{X}$ ثم نضيف ونطرح المتوسطات $\bar{X}_k$ و$\bar{X}_\ell$ و$\bar{X}_\ell$ ونكتب المجموع السابق كما يلي ($\bar{X}_t$ متوسط الطريقة t) :

$$\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\bar{X})^2 = \sum_{k=1}^{g}\sum_{\ell=1}^{g}[(x_{k\ell}-\bar{X}_k-\bar{X}_\ell-\bar{X}_t+2\bar{X})+(\bar{X}_k-\bar{X})+(\bar{X}_\ell-\bar{X})+(\bar{X}_t-\bar{X})]^2 \quad \mathbf{(56-4)}$$

ثم نقوم بتربيع الحدود التي داخل القوس المتوسط، ولكن نظراً لاستقلال العوامل المستخدمة في التحليل فإن مجاميع الجداءات الثنائية تكون معدومة (يمكن البرهان على ذلك لكل حد على حدة . انظر المرجع Dugue (P.282 .

ونتيجة بعض الإصلاحات نحصل على أن مجموع مربعات الانحرافات الإجمالي يساوي :

$$\sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\bar{X})^2 = g\sum_{k=1}^{g}(\bar{X}_k-\bar{X})^2 + g\sum_{\ell=1}^{g}(\bar{X}_\ell-\bar{X})^2 + g\sum_{t=1}^{g}(\bar{X}_t-\bar{X})^2 + \\ + \sum_{k=1}^{g}\sum_{\ell=1}^{g}(x_{k\ell}-\bar{X}_k-\bar{X}_\ell-\bar{X}_t+2\bar{X})^2 \qquad \mathbf{(57-4)}$$

وباستخدام نفس الرموز السابقة نكتب المجاميع السابقة كما يلي :

$$SST = SSA + SSB + SSt + SSE \qquad (58-4)$$

وإن درجات الحرية التي تقابلها تساوي ما يلي :

$$g^2 - 1 = (g-1) + (g-1) + (g-1) + (g-1)\,(g-2) \qquad (59-4)$$

وبعد حساب هذه المجاميع نضعها في جدول تحليل التباين الثلاثي كالتالي :

جدول (4-13): تحليل التباين في المربع اللاتيني :

F المحسوبة	متوسط المربعات	درجة الحرية	مجموع مربعات الانحرافات	مصدر التباين
$F_A = \frac{MSA}{MSE}$	$MSA = \frac{SSA}{g-1}$	$g-1$	$SSA =$	F_A
$F_B = \frac{MSB}{MSE}$	$MSB = \frac{SSB}{g-1}$	$g-1$	$SSB =$	F_B
$F_t = \frac{MSt}{MSE}$	$MSt = \frac{SSt_r}{g-1}$	$g-1$	$SSt_r =$	طرائق المعالجة t_r
	$MSE = \frac{SSE}{(g-1)\,(g-2)}$	$(g-1)\,(g-2)$	$SSE =$	الخطأ العشوائي
	ـــــــ	g^2-1	SST	الاجمالي

أما بالنسبة للفرضيات فهي تنطلق من النموذج (4-54) ونكتبها كما يلي:

$$H_0: \begin{cases} \propto_k = 0 & : K : 1\ 2\ 3 \dots g \\ \beta_\ell = 0 & : \ell : 1\ 2\ 3 \dots g \\ \gamma_t = 0 & : t : 1\ 2\ 3 \dots g \end{cases} \qquad (60-4)$$

أما الفرضية البديلة H_1 فتكتب كما يلي:

$$H_1: \begin{cases} \propto_k \neq 0 & \text{من أجل } k \text{ واحدة على الأقل} \\ \beta_\ell \neq 0 & \text{من أجل} \ell \text{ واحدة على الأقل} \\ \gamma_t \neq 0 & \text{من أجل } t \text{ واحدة على الأقل} \end{cases} \qquad (61-4)$$

وهنا نفترض أن تصميم المربع اللاتيني يعتبر أن تأثيرات الأسطر وتأثيرات الأعمدة ثابتة ، رغم أنها تستخدم للتحكم في مصادر الاختلاف . لذلك فإن تأثيراتها يجب أن تحقق العلاقتين التاليتين :

$$\sum_{k=1}^{g} \propto_k = 0 \qquad \sum_{\ell=1}^{g} \beta_\ell = 0 \qquad (62-4)$$

أما تأثيرات طرائق المعالجة γ_t فإما أن تكون ثابتة أو عشوائية

فإذا كانت γ_t ثابتة فإن تأثيراتها تقدر على أنها انحرافات عن المتوسط العام $\bar{X}$. ولذلك فإنها يجب أن تحقق العلاقة التالية :

$$\sum_{t=1}^{g} \gamma_t = 0 \qquad \left(\text{إذا كانت } \gamma_t \text{ ثابتة}\right) \qquad (63-4)$$

أما إذا كانت γ_t عشوائية فإننا نفترض ونتحقق من أنها خاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$، ثم نتعامل معها بأسلوب آخر (اختبار Tukey) . وهنا نشير إلى أن الفرضية الأساسية التي نريد اختبارها هي فرضية العدم التالية :

$$H_0: \gamma_1 = \gamma_2 = \cdots \gamma_g = 0 \qquad (64-6)$$

مقابل الفرضية البديلة :

من أجل t واحدة على الأقل $H_1: \gamma_t \neq 0$

ولاختبار الفرضية $H_0: \gamma_1 = \gamma_2 = \cdots \gamma_g = 0$ نقوم بحساب مجاميع مربعات الانحرافات ومتوسطاتها ونضعها في جدول تحليل التباين، ونبدأ بحساب مؤشر الاختبار لتأثير المعالجات من العلاقة :

$$F_t = \frac{\frac{SSt}{g-1}}{\frac{SSE}{(g-1)(g-2)}} = \frac{MSt_r}{MSE} \qquad (65-4)$$

ثم نقارن هذه القيمة المحسوبة F_t مع القيمة الحرجة لـ F والمقابلة لمستوى الدلالة $\propto = 0.05$ ولدرجتي الحرية $v_1 = (g-1)$ ، $v_2 = (g-1)(g-2)$ ونتخذ القرار كما يلي :

إذا كانت $F_t \leq F_{v_1\, v_2}(\propto)$ نقبل الفرضية H_0، والتي تقول أن تأثيرات طرائق المعالجة معدومة أو إنها غير معنوية والعكس بالعكس .

أما بالنسبة لاختبار تأثير العاملين F_A و F_B نقوم بحساب قيمتي المؤشرين :

$$F_A = \frac{MSA}{MSE}$$
$$F_B = \frac{MSB}{MSE} \qquad (66-4)$$

ثم نقوم بمقارنة كل منهما مع القيمة الحرجة $F_{v_1\, v_2}(\propto)$، ونقبل الفرضية H_0 حول $\propto_k = 0$ إذا كان $F_A \leq F(\propto)$، كما نقبل الفرضية H_0 حول $\beta_\ell = 0$ إذا كان $F_B \leq F(\propto)$ والعكس بالعكس .

ولكن المؤشرين F_A و F_B يعتبران مؤشرين تقريبيين، لذلك نحسب الكفاءة النسبية للأسطر وللأعمدة من العلاقتين التاليتين :

$$RE\ (\text{للأسطر}) = \frac{MSA + (g-1)MSE}{g * MSE} 100 \qquad (67-4)$$

$$RE\ (\text{للأعمدة}) = \frac{MSB + (g-1)MSE}{g * MSE} 100 \qquad (68-4)$$

ولتقدير الكفاءة الكلية للمربع اللاتيني نستخدم العلاقة :

$$RE\ (\text{للمربع}) = \frac{MSA + MSB + (g-1)MSE}{(g+1)MSE} \qquad (69-4)$$

ملاحظة إذا كانت قيمة إحدى المشاهدات $x_{k\ell}$ مفقودة، فإنه لا يمكننا تطبيق أسلوب المربع للاتيني بدونها، وحتى نتابع العمل علينا أن نقوم بتقديرها من العلاقة :

$$\tilde{x}_{k\ell} = \frac{g(X_k + X_\ell + X_t) - 2X}{(g-1)(g-2)} \qquad (70-4)$$

وهنا علينا أن نخفض درجات الحرية للأخطاء بمقدار درجة واحدة وتصبح كما يلي: $(g-1)(g-2) - 1$

وأن نخفض درجات الحرية لإجمالي المربعات درجة واحدة أيضاً فتصبح : $(g^2 - 2)$

ولتقدير الخطأ المعياري لتقدير المتوسط $\bar{x}_t$ نستخدم العلاقة :

$$S_{\bar{x}_t} = \sqrt{\frac{MSE}{g}} \qquad (71-4)$$

ثم ننشأ مجال الثقة له ذي الاحتمال $(1-\propto)$ من العلاقة:

$$\bar{x}_t \pm t_{(g-1)(g-2)}\left(\frac{\propto}{2}\right) * S_{\bar{x}_k} \qquad (72-4)$$

كما يمكننا حساب الخطأ المعياري للفرق بين متوسطي طريقتين $(\bar{x}_{t1} - \bar{x}_{t2})$ من العلاقة :

$$S_{(\bar{x}_{t1}-\bar{x}_{t2})} = \sqrt{\frac{MSE}{g} + \frac{MSE}{g}} = \sqrt{\frac{2MSE}{g}} \qquad (73-4)$$

ثم إنشاء مجال الثقة له ذي الاحتمال $(1-\propto)$ من العلاقة :

$$(\bar{x}_{t1} - \bar{x}_{t2}) \pm t_{(g-1)(g-2)}\left(\frac{\propto}{2}\right) * S_{(\bar{x}_{t1}-\bar{x}_{t2})} \qquad (74-4)$$

مثال (4-5): لدراسة أداء (6) عمال من عمال شركة النسيج، نرمز لهم بالرموز (A, B, C, D, E, F) تم اعتماد المتحول التابع X الذي يعبر عن إنتاجية العامل في اليوم من القماش (م^2 / يوم) . ولإجراء التجارب تم تجهيز (6) آلات نسيج مختلفة، على أن يتم توزيع هؤلاء العمال على تلك الآلات عشوائياً في كل يوم ولمدة (6) أيام (أيام العمل في الأسبوع)، وبحيث يعمل كل عامل مرة واحدة على آلة مختلفة في كل يوم من ايام العمل في الأسبوع (مرة واحدة في كل عمود ومرة واحدة في كل سطر) .

والمطلوب دراسة تأثير كل من الأيام D والآلات M والعمال L على إنتاجية هؤلاء العمال، إذا علمت أن توزيعات هؤلاء العمال وإنتاجياتهم اليومية كانت بعد تنفيذ الدراسة كما يلي :

جدول (4-14): بيانات المربع اللاتيني للمثال (فرضية) :

الآلات M_ℓ / الأيام D_k	M_1	M_2	M_3	M_4	M_5	M_6	المجموع X_k	المتوسط $\bar{X}_k$
السبت D_1	B 28.7	E 28.4	D 25.4	C 30.7	A 30.6	F 30.9	$X_1 = 174.7$	29.12
الأحد D_2	F 31.4	C 30.1	B 27.4	E 26.8	D 29.8	A 29.8	$X_2 = 175.3$	29.22
الاثنين D_3	D 29.4	F 29.7	A 30.4	B 22.0	E 24.1	C 32.9	$X_4 = 163.0$	28.08
الثلاثاء D_4	A 29.6	B 21.8	E 22.5	F 30.0	C 39.6	D 28.5	$X_3 = 168.5$	27.17
الأربعاء D_5	C 25.8	D 21.9	F 23.1	A 24.3	B 20.7	E 17.7	$X_5 = 133.5$	22.25
الخميس D_6	E 18.1	A 23.6	C 22.5	D 20.2	F 23.7	B 18.9	$X_6 = 127.0$	21.17
المجموع X_ℓ	163.0	155.3	151.3	154.0	159.5	158.7	$X = 942.0$	—
المتوسط $\bar{X}_\ell$	27.17	25.88	25.22	25.67	26.58	26.45	—	$\bar{X} = 26.17$

ولحساب إنتاجية كل عامل على حدة علينا أن تتبع مقدار إنتاجيته على كل آلة وفي كل يوم . لذلك نتبع إنتاجية هؤلاء العمال حسب الآلات ونصمم جدولاً خاصاً لتبويبها من جديد حسب العمال والآلات فنحصل من الجدول (4−10) السابق على الجدول التالي :

جدول (4−15) إنتاجية العمال حسب الآلات :

العمال / الآلات	A	B	C	D	E	F	المجموع X_ℓ	المتوسط $\bar{X}_\ell$
M_1	29.6	28.7	25.8	29.4	18.1	31.4	163.0	27.17
M_2	23.6	21.8	30.1	21.9	28.4	29.7	155.5	25.88
M_3	30.4	27.4	22.5	25.4	22.5	23.1	151.3	25.22
M_4	24.3	22.0	30.7	20.2	26.8	30.0	154.0	25.67
M_5	30.6	20.7	30.6	29.8	24.1	23.7	159.5	26.58
M_6	29.8	18.9	32.9	28.5	19.7	30.9	158.7	26.45
X_t	168.3	139.5	172.6	155.2	137.7	168.8	942.0	—
$\bar{X}_t$	28.05	23.25	28.77	25.87	22.95	28.13	—	26.17

ومن السطر الأخير لهذا الجدول نلاحظ أن إنتاجية هؤلاء العمال تختلف من عامل لآخر . وإن أحسنها هي إنتاجية العمال C ثم F ثم A . وإن أسوأها هي إنتاجية العمال E ثم B ثم D . وكذلك يمكننا أن نبوب إنتاجية هؤلاء العمال حسب الأيام فنحصل من الجدول (4−14) على جدول مشابه للجدول (4−15) السابق . إلا أن حساب الجدول (4−15) يغني عنه في الحسابات اللاحقة . والآن نعتمد النموذج النظري التالي :

$$x_{k\ell} = \bar{X} \pm \propto_k + B_\ell + \gamma_t + e_{k\ell}$$

ثم نضع الفرضيات كما يلي:

$$H_0: \begin{cases} \propto_1 = \propto_2 = \cdots = \propto_6 = 0 \\ B_1 = B_2 = \cdots = B_6 = 0 \\ \gamma_1 = \gamma_2 = \cdots = \gamma_6 = 0 \end{cases}$$

$$H_1: \begin{cases} \propto_k \neq 0 & \text{من أجل } k \text{ واحدة على الأقل} \\ B_\ell \neq 0 & \text{من أجل } \ell \text{ واحدة على الأقل} \\ \gamma_t \neq 0 & \text{من أجل } t \text{ واحدة على الأقل} \end{cases}$$

ولاختبار هذه الفرضيات نقوم بحساب مجاميع مربعات الانحرافات SST وSSA وSSB وSSt وSSE من بيانات الجدولين (4−14) و(4−15) ونستخدم العلاقات التالية :

$$SST = \sum_{k=1}^{6}\sum_{\ell=1}^{6}(x_{k\ell} - \bar{X})^2 = \sum_{k=1}^{6}\sum_{\ell=1}^{6} x_{k\ell}^2 - \frac{X^2}{g^2}$$

$$SST = [(28.7)^2 + (28.4)^2 + \cdots (23.7)^2 + (18.9)^2] - \frac{(942)^2}{36}$$

$$SST = 25299.10 - 24649 = 650.10$$

ثم نقوم بحساب مجموع المربعات للأيام D فنجد من مجاميع الأسطر في الجدول (4-14) أن:

$$SSA = \sum_{k=1}^{6} X_k^2 - \frac{X^2}{g^2} = [(174.7)^2 + (175.3)^2 + \cdots (127.0)^2] - \frac{(942)^2}{36} = 378.11$$

ثم نقوم بحساب مجموع المربعات للآلات فنجد من مجاميع الأعمدة في الجدول (4-14) أن:

$$SSB = \sum_{\ell=1}^{6} X_\ell^2 - \frac{X^2}{g^2} = [(163.3)^2 + (155.5)^2 + \cdots (158.7)^2] - \frac{(942)^2}{36} = 14.81$$

ثم نقوم بحساب مجموع المربعات للعمال اعتماداً على أعمدة الجدول (4-15) فنجد أن:

$$SSt = \sum_{t=1}^{6} X_t^2 - \frac{X^2}{g^2} = [(168.3)^2 + (139.5)^2 + \cdots (168.8)^2] - \frac{(942)^2}{36} = 199.36$$

ثم نقوم بحساب SSE من العلاقة :

$$SSE = SST - SSA - SSB - SSt$$

$$SSE = 650.10 - 378.4 - 14.81 - 199.36 = 57.82$$

وأخيراً نضع نتائج هذه الحسابات مع درجات حرياتها في جدول تحليل التباين الثلاثي فنحصل على أن :

جدول (4-16): تحليل التباين للمربع اللاتيني :

قيمة F المحسوبة	متوسط المربعات	درجة حرية	مجموع المربعات	مصدر التباين
$F_A = 26.16$	75.62	$g - 1 = 5$	$SSA = 378.11$	الأسطر (الأيام)
$F_B = 1.02$	2.96	$g - 1 = 5$	$SSB = 14.8$	الأعمدة (الآلات)
$F_t = 13.79$	39.87	$g - 1 = 5$	$SSt = 199.36$	المعالجات (العمال)
ـــــــ	2.89	$(g-1)(g-2) = 20$	$SSE = 57.82$	الأخطاء
ـــــــ	ـــــــ	$g^2 - 1 = 35$	$SST = 650.10$	الإجمالي

ثم نقوم بإيجاد قيمة F الحرجة الموافقة لدرجتي حرية $v_1 = 5$, $v_2 = 20$ ولمستوى دلالة $\propto = 0.05$ فنجد أن :

$$F_{v_1,v_2}(\propto) = F_{5,20}(0.05) = 2.71$$

وهي نفسها لجميع العوامل، لذلك نقارن F_A وF_B وF_t معها، فنجد ما يلي :

1- بالنسبة للعمال نجد أن: $F_t > F(\propto)$، لذلك نرفض فرضية العدم $H_0: \gamma_1 = \gamma_2 = \cdots = \gamma_6 = 0$ ونقبل الفرضية البديلة H_1 . التي تقول أن الإنتاجية X تتأثر بالعمال (وهذا أمر واضح من السطر الأخير في الجدول (6-15) .

2- بالنسبة للأيام (الأسطر) نجد أن $F_A > F(\propto)$، لذلك نرفض فرضية العدم $H_0: \propto_1 = \propto_2 = \cdots = \propto_6 = 0$ ونقبل الفرضية البديلة H_1. التي تقول أن الإنتاجية X تتأثر بالأيام (وهذا أمر واضح من العمود الأخير في الجدول (6-10) .

3- بالنسبة للآلات (الأعمدة) نجد أن $F_B < F(\propto)$ لذلك نقبل فرضية العدم $H_0: B_1 = B_2 = \cdots = B_6 = 0$ ونستنتج أن الآلات لا تؤثر على إنتاجية العمال .

ولتقدير كفاءة هذا التصميم نقوم بحساب الكفاءة النسبية للمربع ككل فنجد أن :

$$RE(Squere) = \frac{MSA + MSB + (g-1)MSE}{(g+1)MSE}100 = \frac{75.62 + 2.96 + 5 * 2.89}{7 * 2.89}100 = 459.86\%$$

وهذا يدل على كفاءة عالية بالنسبة للقطاعات العشوائية التامة .

كما يمكننا حساب الكفاءة النسبية للأسطر (بدون الأعمدة) من العلاقة :

$$RE(row) = \frac{MSA + (g-1)MSE}{g * MSE}100 = \frac{75.62 + 5 * 2.89}{6 * 2.89}100 = 519.43\%$$

وهذا يدل على كفاءة عالية بالنسبة للقطاعات العشوائية التامة .

وكذلك نقوم بحساب الكفاءة النسبية للأعمدة (بدون الأسطر) من العلاقة :

$$RE(column) = \frac{MSB + (g-1)MSE}{g * MSE}100 = \frac{2.96 + 5 * 2.89}{6 * 2.89}100 = 100.40\%$$

وهذا يدل على أن كفاءة الأعمدة (الآلات) بقيت تساوي نفسها 100% ولم تؤثر على إنتاجية العمال .

4-5: تحليل التباين المشترك (تحليل التغايرANCOVA):

4-5-1: تمهيد:

لقد استعرضنا في النماذج السابقة لتحليل التباين النماذج التي تدرس تغيرات متحول تابع X (Dependent variable)، الناتجة عن متغير أو متغيرات وصفية مستقلة (Independent)، مثل المجتمعات أو المعالجات أو القطاعات .

ولكن إذا كان التابع X مرتبطاً بمتحول كمي آخر Y غير مرغوب به، ولكنه ملازم لـ X ولا نستطيع التحكم فيه، ولكننا نريد التخلص من تأثيره على التابع X . فإننا نلجأ إلى إجراء تحليل التباين المشترك (التغاير ANCOVA) لفصل تأثيرات Y غير المرغوب فيها عن تغيرات X ونتبع في ذلك إحدى الطريقتين :

1- إزالة اختلافات X المرتبطة بـ Y من قياسات X (إذا كان ذلك ممكناً) وعندها نحصل على قياسات صافية لـ X . ثم نقوم بإجراء تحليل التباين على القياسات الصافية فنحصل على اختبارات قوية (ولكن هذه الطريقة قد لا تكون ممكنة في معظم الحالات) .

2- تعديل متوسطات المعالجات في المجتمعات، بحيث يتم طرح قيم موحدة للمتغير المستقل Y منها، وبالتالي نحصل على طريقة عادلة لمقارنة قيم متوسطات X في تلك المجتمعات المختلفة .

فمثلاً لدراسة تغيرات الانفاق الشهري X لطلاب الجامعة حسب الجنس (مجتمع الذكور ومجتمع الإناث). نلاحظ أن ذلك الانفاق لا يتأثر بنوع الجنس فقط، بل يتأثر أيضاً بمقدار الدخل الشهري المخصص للطالب والذي سنرمز له بـ Y . لذلك علينا أن نقوم بعزل تأثير الدخل Y على X . وذلك حتى نتمكن من إجراء مقارنة عادلة لمتوسطي الإنفاق X حسب الجنس دون تأثير الدخل Y عليهما .

وللتخلص من تأثيرات المتحول الإضافي Y على قيم X نلجأ إلى تحليل التباين المشترك (التغاير ANCOVA)، وإذا اجرينا تحليل التباين البسيط (ANOVA) بدون عزل تأثير Y، فإن ذلك سيؤدي إلى

تضخيم الخطأ التجريبي، ويصبح من الصعب اكتشاف الفروقات الحقيقية بين المجتمعات . وبذلك نجد أن المهمة الرئيسية لتحليل التغاير هي تصغير قيمة الخطأ التجريبي .

وتتضمن المراجع المختصة عدة أنواع لتحليل التغاير هي:

- تحليل التغاير في تصميم العشوائية التامة (CRD) .
- تحليل التغاير في تصميم القطاعات العشوائية (RCBD) .
- تحليل التغاير في تصميم التجارب العاملية 2^k $(2^k FD)$.

وسنقتصر في منشورنا هذا على النوع الأول لأنه الأكثر انتشاراً والأسهل تطبيقاً .

4-5-2: تحليل التغاير في تصميم العشوائية التامة (CRD) باتجاه واحد (ANCOVA one way) .

يمكن النظر إلى تحليل التغاير على أنه تعديل لتحليل التباين البسيط باتجاه واحد (ANOVA)، وذلك بعد إضافة المتحول Y إلى النموذج الرياضي ثم إعادة صياغة النموذج بحيث يتم عزل تأثير Y منه . وبصورة عامة نفترض أنه لدينا g مجتمعاً، تؤثر على متحول تابع X ، وأن X يترافق مع متحول إضافي كمي Y .

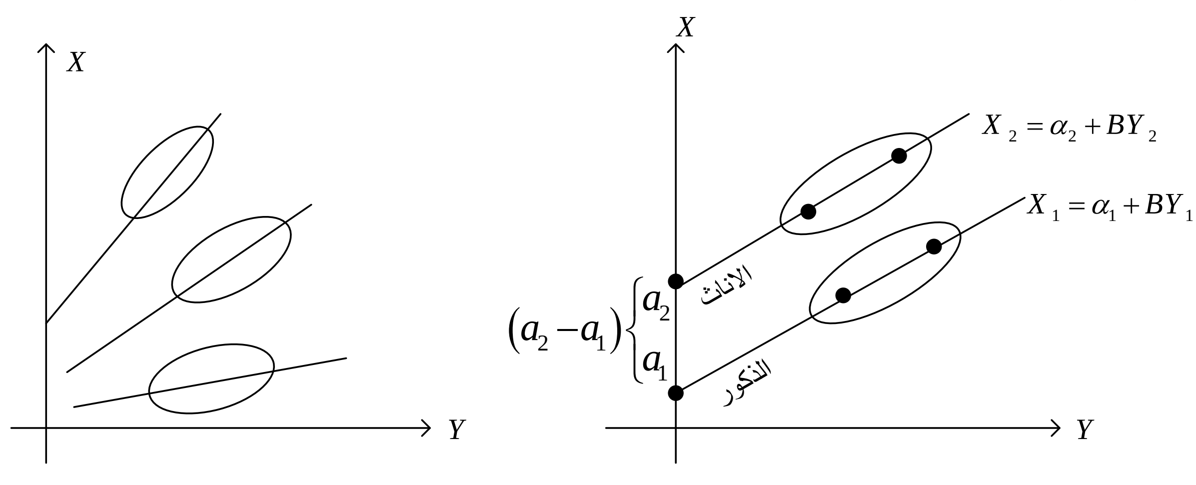
وأن X يرتبط مع Y في كل مجتمع k بعلاقة انحدار خطية من الشكل:

$$X_k = a_k + \beta_k Y_k = a_k + \beta Y_k \qquad : k: 1\ 2\ 3 \dots g \qquad (75-4)$$

ولتسهيل صياغة النموذج الرياضي المشترك نفترض أن قيم الأمثال β_k متساوية في جميع تلك المجتمعات، أي إننا نفترض أو نشترط أن يكون :

$$\beta_1 = \beta_2 \dots = \beta_g = \beta \qquad (75\,a-4)$$

وإن هذا يعني أن ميول مستقيمات هذه العلاقات متساوية في جميع المجتمعات المدروسة، وترسم الشكل اليميني التالي :



حالة الأمثال غير متساوية

حالة الأمثال المتساوية $\beta_1 = \beta_2 = \beta$

الشكل (4-3): حالات علاقة x بـ Y

كما نفترض أن حجوم العينات المسحوبة من تلك المجتمعات متساوية وتساوي $n_k = r$،ومن جهة ثانية نجد أنه يمكننا صياغة نموذج تحليل التباين البسيط (ANOVA) للمتحول التابع X حسب (4–7) كما يلي:

$$x_{ki} = \mu_x + \propto'_k + e'_{ki} \qquad (76-4)$$

حيث أن: $g ... 3\ 2\ 1 = k$ و $r ... 3\ 2\ 1 = i$

وحيث أن: μ_x هو توقع X، وإن: $\propto'_k$ هو مقدار تأثير المجتمع k على X، وهو يحقق الشرط التالي: $\sum_{k=1}^{g} \propto'_k = 0$.

وأن: e'_{ki} هي حدود الخطأ العشوائي (البواقي) ويشترط فيها أن تكون مستقلة وخاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$.

ومن جهة ثالثة نجد أنه يمكننا صياغة النموذج الرياضي لتحليل التباين (ANOVA) للمتحول الإضافي Y حسب (4–7) كما يلي:

$$y_{ki} = \mu_y + \propto''_k + e''_{ki} \qquad (77-4)$$

حيث أن: $g ... 3\ 2\ 1 = k$ و $r ... 3\ 2\ 1 = i$

وحيث أن: μ_y هو توقع Y، وإن: $\propto''_k$ هو مقدار تأثير المجتمع k على Y، وهو يحقق الشرط التالي: $\sum_{k=1}^{g} \propto''_k = 0$.

وأن: e''_{ki} هي حدود الخطأ العشوائي (البواقي) ويشترط فيها أن تكون مستقلة وخاضعة للتوزيع الطبيعي $N(0\,,\sigma^2)$.

وبعدها نضرب طرفي العلاقة (4–77) بقيمة الأمثال β (بعد حسابها من(4–75))، ونطرح أطرافها من أطراف العلاقة (4–76) فنحصل على ما يلي:

$$(x_{ki} - \beta y_{ki}) = (\mu_x - \beta\mu_y) + (\propto'_k - \beta \propto''_k) + (e'_{ki} - \beta e''_{ki}) \qquad (78-4)$$

وإذا رمزنا للطرف الأيسر بـ $Z_{ki} = (x_{ki} - \beta y_{ki})$ وأخذنا توقعه نجد أن:

$$E(Z_{ki}) = E(x_{ki}) - \beta E(y_{ki}) = \mu_x - \beta\mu_y = \mu_z \qquad (79-4)$$

وهكذا يظهر لدينا متحول جديد هو: $Z = X - \beta Y$

وأن توقعه: $\mu_z = \mu_x - \beta\mu_y$

وإذا رمزنا للحدود الأخرى في (4–78) بالرموز التالية:

$\propto_k = \propto'_k - \beta \propto''_k$ و $e_{ki} = e'_{ki} - \beta e''_{ki}$ فإننا نحصل على النموذج التالي:

$$Z_{ki} = \mu_z + \propto_k + e_{ki} \qquad (80-4)$$

حيث أن: $g ... 3\ 2\ 1 = k$ و $r ... 3\ 2\ 1 = i$ وهو نموذج تحليل التباين البسيط لـ Z .

حيث أن: μ_z هو توقع المتحول Z، وأن: $\propto_k$ هو تأثير المجتمع k على Z وهو يحقق الشرط التالي: $\sum_{k=1}^{g} \propto_k = 0$.

وأن e_{ki}: هي حدود الخطأ العشوائي الجديدة . وهي حدود مستقلة وتخضع للتوزيع الطبيعي $N(0\,,\sigma^2)$.

وهكذا نكون قد توصلنا إلى نفس الصيغة الرياضية التي يأخذها نموذج ANOVA للمتحول الجديد Z.

ولإظهار المتحول Y في النموذج (4−80) نعيد الرموز إلى أصولها وندمج الحدود المتشابهة. فنجد أن العلاقة (4−80) تأخذ الشكل التالي:

$$(x_{ki} - \beta y_{ki}) = (\mu_x - \beta\mu_y) + \propto_k + e_{ki}$$

وبعد الإصلاح نحصل على الصيغة التي تعطينا أية قيمة x_{ki} كما يلي:

$$x_{ki} = \mu_x + \propto_k + \beta(y_{ki} - \mu_y) + e_{ki} \qquad (81-4)$$

حيث أن: $k = 1\ 2\ 3 \dots g$ و $i = 1\ 2\ 3 \dots r$.

وهي صيغة نموذج تحليل التغاير ANCOVA للمتحول X المرتبط بالمتحول الإضافي Y. وهي الصيغة الأكثر انتشاراً .

ويمكننا الحصول على صيغة أخرى لـ (4−81) إذا قمنا بكتابتها كما يلي:

$$x_{ki} = (\mu_x + \beta\mu_y) + \propto_k + \beta y_{ki} + e_{ki}$$

وإذا رمزنا للمقدار $\mu' = \mu_x + \beta\mu_y$ نحصل على الصيغة التالية :

$$x_{ki} = \mu' + \propto_k + \beta y_{ki} + e_{ki} \qquad (82-4)$$

حيث أن: $k = 1\ 2\ 3 \dots g$ و $i = 1\ 2\ 3 \dots r$ ، وحيث أن: μ' هو مقدار ثابت يعبر عن توقع متحول ثالث وليس عن توقع X، لأن توقع X يساوي $\mu_x = \mu' + \beta\mu_y$.

وهكذا نكون قد توصلنا إلى نموذج تحليل التغاير للمتحول X المرتبط بالمتحول الإضافي Y، وهو يأخذ إحدى الصيغتين المتكافئتين (4−81) و(4−82) .

وأخيراً نشير إلى أنه لتطبيق هذا النموذج يشترط على عناصره أن تحقق افتراضات تحليل التباين ANOVA وافتراضات الانحدار البسيط ونلخصها بما يلي:

- الافتراضات على نموذج ANCOVA :

1− أن تكون العينات المسحوبة عشوائية ومستقلة، وأن تكون حجومها متساوية وتساوي: $n_1 = n_2 = \dots = n_g = r$.

2−أن يكون X وY متحولان كميان وخاضعان في كل مجتمع k للتوزيع الطبيعي بتوقعين μ_{kx} و μ_{ky} على الترتيب .

3−أن تكون تباينات التابع X في جميع المجتمعات متساوية وتساوي σ^2 .

4−أن تكون قياسات المتحول الإضافي Y نهائية، ولا تتضمن أخطاءً في القياس ولا تتأثر بالمجتمعات أو المعالجات .

5−أن تكون العلاقات بين X وY في جميع المجتمعات خطية وتأخذ الشكل التالي: $X_k = a_k + \beta_k Y_k$ وتكون $\beta_k \neq 0$.

6−أن تكون قيم جميع الأمثال β_k متساوية في جميع المجتمعات المدروسة أي $\beta_k = \beta$، أي أن تكون المستقيمات متوازية كما في الشكل (4−3) السابق .

7−أن يكون مجموع تأثيرات المجتمعات على X معدوماً . أي أن يكون: $\sum_{k=1}^{g} \propto_k = 0$.

8–أن تكون الأخطاء e_{ki} أو البواقي الناتجة عن النموذج مستقلة وخاضعة للتوزيع الطبيعي بتوقع معدوم وتباين ثابت σ^2 أي لـ $N(0\,,\sigma^2)$.

ونلاحظ من هذه الافتراضات أن الافتراض الرابع ينص على أن المتغير الإضافي Y لا يتأثر بالمجتمعات (المعالجات) وهو أهم افتراض في تحليل التغاير ، لأنه إذا كان Y يتأثر بالمجتمعات، فإن معاملات الانحدار β_k ستختلف من مجتمع لآخر، وهذا يؤدي إلى عدم ثبات قيم تلك المعاملات في المجتمعات المختلفة، وهذا يخل في الافتراض السادس $(\beta_k = \beta)$ ويجعل النموذج (4–81) غير صالح للتطبيق على تلك المجتمعات.

ونستنتج مما سبق عند تطبيق ANCOVA أنه يجب علينا قبل كل شيء التحقق من أن Y لا يتأثر بالمجتمعات، وذلك بإجراء تحليل التباين ANOVA على Y بمفرده، فإذا كانت النتيجة قبول فرضية العدم H_0 (لا توجد فروقات لـ Y بين المجتمعات)، فإننا نتابع العمل للتحقق من الافتراضين الخامس $(\beta_k \neq 0)$ والسادس $(\beta_k = \beta)$, ولذلك يجب علينا أن نقوم بإيجاد معاملات العلاقات الخطية بين X Y في جميع المجتمعات $X_k = a_k + \beta_k Y_k$ بطريقة المربعات الصغرى، ثم القيام باختبار عدم وجود فروقات معنوية بين المعاملات β_k وذلك بوضع الفرضيتين كما يلي:

$H_0 :\ \beta_k =\ \beta_0 \qquad : k = 1\ 2\ 3 \dots g$

$H_1 :\ \beta_k \neq \beta_0$ على الأقل من أجل قيمة واحدة لـ k

حيث أن β_0 هي قيمة افتراضية يضعها الباحث أو يحسبها من متوسط المعاملات β_k، وللتحقق من H_0 نقوم بحساب قيمة اختبار (ستودينت) t المعرف بالعلاقة :

$$t_k = \frac{\beta_k -\ \beta_0}{S(\beta_k)} \qquad : k = 1\ 2\ 3 \dots g \qquad (83-4)$$

حيث $S(\beta_k)$ هو الانحراف المعياري للمعامل β_k .

ثم نقوم بمقارنة قيمة t_k المحسوبة مع قيمة t الحرجة والمقابلة لـ $(n-1)$ درجة حرية ولنصف مستوى الدلالة $\left(\frac{\propto}{2}\right)$.

فإذا كانت $|t_k| \leq t_{n-1}\left(\frac{\propto}{2}\right)$ فإننا نقبل H_0 ونعتبر المعاملات β_k متساوية باحتمال 0.95 وبذلك يكون الافتراض السادس محققاً .

ولاختبار تحقق الافتراض الخامس نطبق الاختبار t على الفرضيتين التاليتين كما يلي:

$H_0 :\ \beta_0 =\ 0 \qquad H_1 :\ \beta_0 \neq 0$

ونطبق نفس المؤشر المعرف في (6–83) على جميع β_k بطريقة إعادة الاختبار. فإذا كانت النتيجة هي رفض H_0 من أجل جميع β_k، فإننا نعتبر أن $\beta_0 \neq 0$ وأنه يوجد علاقة خطية معنوية بين X وY في تلك المجتمعات، ويمكن التحقق من الافتراض الخامس عن طريق معامل الارتباط r أو معامل التحديد R^2 ...الخ.

وبناءً على ذلك يتم إجراء تحليل التغاير ANCOVA للتابع X المرتبط بمتحول إضافي Y. وذلك ضمن تحقق الافتراضات المذكورة أعلاه .

وبعد هذه المقدمة ننتقل إلى إجراء التحليل اللازم للنموذج (4-81) ونضع الفرضيتين الخاصتين به (العدم والبديلة) على الشكل التالي :

$$H_0 : \propto_1 = \propto_2 = \propto_3 = \cdots = \propto_g = 0$$

$$H_1 : \propto_k \neq 0 \quad \text{من أجل } k \text{ واحدة على الأقل} \qquad (84-6)$$

ولاختبار الفرضيات (4-84) علينا أن نقوم بحساب مجاميع مربعات الانحرافات لكل من المتحولين X وY ولجدائهما $X * Y$, وذلك باستخدام العلاقات الرياضية التالية (مع الانتباه هنا إلى أن r هو الحجم الموحد للعينات المسحوبة من تلك المجتمعات و g عدد المجتمعات) :

بالنسبة للمتحول X نجد أن:

$$SST_x = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X})^2 = \sum_{k=1}^{g}\sum_{i=1}^{r} x_{ki}^2 - \frac{X^2}{g * r} \qquad (85-4)$$

حيث أن: $X = \sum_{k=1}^{g}\sum_{i=1}^{r} x_{ki}$ هو مجموع قيم X في جميع العينات .

$$SSA_x = \sum_{k=1}^{g}(\bar{X}_k - \bar{X})^2 = \sum_{k=1}^{g}\frac{X_k^2}{r} - \frac{X^2}{g * r} \qquad (86-4)$$

حيث أن: $X_k = \sum_{i=1}^{r} x_{ki}$ هو مجموع قيم X في العينة k فقط و$\bar{X}_k$ متوسطها .

$$SSE_x = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X}_k)^2 = SST_x - SSA_x \qquad (87-4)$$

ولها درجات الحرية : $g(r-1) \ = (gr-1) \ \ (g-1)$

أما بالنسبة للمتحول Y فنجد أن:

$$SST_y = \sum_{k=1}^{g}\sum_{i=1}^{r}(y_{ki} - \bar{Y}) = \sum_{k=1}^{g}\sum_{i=1}^{r} y_{ki}^2 - \frac{Y^2}{g * r} \qquad (88-4)$$

حيث أن: $Y = \sum\sum y_{ki}$ هو مجموع قيم Y في جميع العينات .

$$SSA_y = \sum_{k=1}^{g}(\bar{Y}_k - \bar{Y})^2 = \sum_{k=1}^{g}\frac{Y_k^2}{r} - \frac{Y^2}{g * r} \qquad (89-4)$$

حيث أن: $Y_k = \sum_{i=1}^{r} y_{ki}$ هو مجموع قيم Y في العينة k فقط ومتوسطها $\bar{Y}_k$.

$$SSE_y = \sum_{k=1}^{g}\sum_{i=1}^{r}(y_{ki} - \bar{Y}_k)^2 = SST_y - SSA_y \qquad (90-4)$$

أما بالنسبة لمجاميع الجداء $X * Y$ فنجد أن:

$$SPT_{xy} = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X})(y_{ki} - \bar{Y}) = \sum_{k=1}^{g}\sum_{i=1}^{r} x_{ki} * y_{ki} - \frac{X * Y}{gr} \qquad (91-4)$$

حيث أن: $X = \sum\sum x_{ki}$ وأن $Y = \sum\sum y_{ki}$ وأن درجة حريته $(gr-1)$.

$$SPE_{xy} = \sum_{k=1}^{g}\sum_{i=1}^{r}(x_{ki} - \bar{X}_k)(y_{ki} - \bar{Y}_k) = \sum_{k=1}^{g}\left[\sum_{i=1}^{r}(x_{ki} * y_{ki}) - \frac{X_k * Y_k}{gr}\right] = \sum_{k=1}^{g} SPE_k \qquad \mathbf{(92-4)}$$

حيث أن: $X_k = \sum_{i=1}^{r} x_{ki}$ وأن: $Y_k = \sum_{i=1}^{r} y_{ki}$ وأن درجة حريته $= g(r-1)$.

$$SPA_{xy} = \sum_{k=1}^{g} (\bar{X}_k - \bar{X})(\bar{Y}_k - \bar{Y}) = \sum_{k=1}^{g} \frac{X_k Y_k}{r} - \frac{XY}{gr} = SPT_{xy} - SPE_{xy} \qquad (\mathbf{93-4})$$

وهنا نشير إلى أن المجاميع الجدائية SPT_{xy} و SPE_{xy} SPA_{xy} يمكن أن تكون سالبة على عكس مجاميع المربعات لـ X و Y الموجبة دائماً , ولسهولة العرض نضع هذه المربعات والجداءات في جدول منظم كالتالي:

جدول (4-17): جدول التحليل قبل التعديل

مصدر التباين	درجة الحرية dk	مربعات المتحول X	مربعات المتحول Y	مجاميع الجداء $X * Y$
المجتمعات (المعالجات) SSA	$g-1$	SSA_x	SSA_y	SPA_{xy}
الخطأ العشوائي SSE	$g(r-1)$	SSE_x	SSE_y	SPE_{xy}
الاجمالي SST	$g\,r-1$	SST_x	SST_y	SPT_{xy}

وأخيراً نقوم بحساب مجاميع مربعات الانحرافات المعدلة بعد عزل تأثير المتحول الإضافي Y، من العلاقات التالية:

$$(SST)' = (SST)_x - \frac{(SPT)_{xy}^2}{(SST)_y} \qquad : df = gr-2 \qquad (94-4)$$

$$(SSE)' = (SSE)_x - \frac{(SPE)_{xy}^2}{(SSE)_y} \qquad : df = g(r-1)-1 \qquad (95-4)$$

$$(SSA)' = (SST)' - (SSE)' \qquad : df = g-1 \qquad (96-4)$$

وبناء على ذلك يمكننا أن نضع هذه النتائج في جدول تحليل التباين المشترك المعدل والذي يأخذ الشكل(4-18) التالي، ومنه نلاحظ أنه لقد تم تخفيض درجة حرية الاجمالي ودرجة حرية الخطأ التجريبي بمقدار درجة واحدة عما كانت عليه في الجدول (4-17)، وذلك لأننا استخدمنا بيانات العينة في تقدير β والذي يساوي: $\tilde{\beta} = \frac{(SSE)_{xy}^2}{SSE_y}$.

أما المقدار : $\frac{(SPT)_{xy}^2}{SST_y}$ فهو مجموع مربعات الانحدار للنموذج (4-81) بدون اعتبار تأثير المجتمعات . والذي يأخذ الشكل التالي (بدون $\propto_k$) :

$$x_{ki} = \mu_x + \beta(y_{ki} - \bar{Y}) + e_{ki}$$

وهو يعبر عن كمية التباين في قيمة X الناجمة عن المتحول المستقل Y .

جدول (4-18): جدول تحليل التباين المشترك المعدل:جدول ANCOVA

مصدر التباين	الرمز	درجة حرية	متوسط مجموع المربعات المعدلة	قيمة F المحسوبة
بين المجتمعات أو المعالجات	$(SSA)'$	$g-1$	$MSA' = \frac{SSA'}{g-1}$	$F = \frac{MSA'}{MSE'}$
الأخطاء العشوائية	$(SSE)'$	$g(r-1)-1$	$MSE' = \frac{SSE'}{g(r-1)-1}$	———
الاجمالي (المعدل)	$(SST)'$	$g * r - 2$	———	———
	حيث أن : r هو حجم العينة الموحد في المجتمعات			

ثم نقوم بإجراء الاختبار اللازم باستخدام المؤشر F المعرف بالعلاقة التالية :

$$F = \frac{\frac{SSA'}{g-1}}{\frac{SSE'}{(g(r-1)-1)}} = \frac{MSA'}{MSE'} \sim F_{(g-1),(g(r-1)-1)} \qquad (97-4)$$

وهو يخضع لتوزيع F بدرجتي حرية $v_1 = (g-1)$ و$v_2 = (g(r-1)-1)$، لذلك نقارن هذه القيمة المحسوبة مع القيمة الحرجة للمتحول F عند مستوى دلالة $\propto$ ونتخذ القرار كما يلي:

إذا كانت $F \le F_{v_1,v_2}(\propto)$ نقبل فرضية العدم H_0

أما إذا كانت $F > F_{v_1,v_2}(\propto)$ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 بمستوى دلالة $\propto$.

ملاحظة: لا يجوز إجراء تحليل التباين المشترك ANCOVA , قبل حساب معادلة الانحدار الخطي بين X وY واختبار معنوية β , والتي سيكون لها الصيغة التالية :

$$X = a + \beta Y \qquad (98-4)$$

فإذا كانت قيمة β غير معنوية (معدومة $\beta = 0$ في جميع المجتمعات) فإن ذلك يعني أن X غير مرتبط بـ Y ، ولا داعي لإدخال Y في النموذج وتعديله، وعندها نقوم بحساب تحليل التباين البسيط باتجاه واحد كالعادة على X فقط .

أما إذا كانت قيمة β معنوية (غير معدومة), فإننا نختبر قيمها إذا كانت متساوية في جميع المجتمعات أم لا.

فإذا كانت قيم β متساوية في جميع المجتمعات نقوم بإجراء تحليل ANCOVA ونعدل النموذج كما سبق.

أما إذا كانت قيم β غير متساوية في المجتمعات المدروسة , فإننا نقوم بحساب معادلات الانحدار في كل مجتمع على حدة ونتخذ القرار المناسب حول الأسلوب المناسب لتفسير أسباب تغيرات X .

ملاحظة 2: يمكن تعميم أسلوب تحليل ANCOVA على متحولين إضافيين أو أكثر Y_1 وY_2 مرتبطين بالمتحول التابع المدروس X، وذلك باتباع نفس المعالجة وباستخدام نفس الفرضيات والاختبارات .

مثال (4-6): لنفترض أننا نريد دراسة تغيرات علامة الرياضيات للطلاب الدارسين في 3 مدارس محددة (3 مجتمعات) . لذلك نرمز لعلاماتهم في مقرر الرياضيات فيها بـ X، ونريد أن ندرس تغيرات X الناتجة على اختلاف تلك المدارس(أي دراسة تأثير المدارس على علامة الرياضيات) . إلا أن أحد المختصين لفت انتباهنا إلى أن علامة الرياضيات X مرتبطة أيضاً بدرجات ذكاء الطالب Y , لذلك يجب عزل تأثيره. ولهذا علينا استخدام أسلوب تحليل ANCOVA وفق النموذج المعدل (4-81) التالي :

$$x_{ki} = \mu_x + \propto_k + \beta(y_{ki} - \mu_y) + e_{ki} \qquad (99-4)$$

ولذلك قمنا بسحب (3) عينات عشوائية من طلاب هذه المدارس بحجم موحد $n_k = r = 10$ ، وأخذنا منهم علامات الرياضيات وحددنا درجة الذكاء لكل طالب، فكانت كما في الجدول التالي (العلامات X والدرجة Y حسبت من 100) .

جدول (4–19): العلامات X والدرجات Y مع مجاميعهما

رقم الطالب	المدرسة الاولى		المدرسة الثانية		المدرسة الثالثة		المجموع	
	Y_{1i}	X_{1i}	Y_{2i}	X_{2i}	Y_{3i}	X_{3i}	Y	X
1	73	55	50	76	82	62		
2	60	70	66	80	88	90		
3	45	30	90	86	90	82		
4	33	27	86	70	50	40		
5	90	89	91	85	70	42		
6	68	50	80	73	75	80		
7	77	60	50	40	80	90		
8	80	98	40	35	95	60		
9	85	79	47	25	40	30		
10	70	82	90	60	50	44		
$\sum$ المجموع	680	640	690	630	720	620	$Y = 2090$	$X = 1890$
معادلات الانحدار	$\tilde{X}_1 = -12.87 + 1.130Y$ $r_1 = 0.836$		$\tilde{X}_2 = 10.42 + 0.762Y_2$ $r_2 = 0.712$		$\tilde{X}_3 = -3.54 + 0.910Y_3$ $r_3 = 0.7745$			

ولدراسة تغيرات X بمعزل عن Y نقوم أولاً بحساب معادلات الانحدار الخطي لعلاقة X بـ Y في كل مدرسة على حدة فنجد أنها تساوي (انظر السطر الأخير من الجدول):

$$\tilde{X}_1 = a + bY_1 = -12.87 + 1.130\ Y_1 \quad (r_1 = 0.836)$$
$$\tilde{X}_2 = a + bY_2 = 10.42 + 0.762\ Y_2 \quad (r_2 = 0.712)$$
$$\tilde{X}_3 = a + bY_3 = -3.54 + 0.910\ Y_3 \quad (r_3 = 0.7745)$$

وبدراسة هذه المعادلات ومعاملات الارتباط فيها نلاحظ أن قيم b فيها ليست معدومة (وليست قريبة من الصفر) . ولذلك فإننا نتابع التحليل ANCOVA دون الخوض في مسألة البرهان على ذلك . ونترك مسألة البرهان على تساوي أو عدم تساوي قيم b في المجتمعات الثلاثة إلى القارئ على سبيل التمرين .

ولمتابعة تحليل ANCOVA نضع النموذج الرياضي كما يلي:

$$x_{ki} = \mu_x + \propto_k + \beta\left(y_{ki} - \mu_y\right) + e_{ki}$$

ثم نضع الفرضيتين كما يلي:

$$H_0 : \propto_1 = \propto_2 = \propto_3 = \cdots = \propto_g = 0 \qquad \left(\text{بشرط } \sum \propto_k = 0\right)$$

$$H_1 : \propto_k \neq 0$$ من أجل k واحدة على الأقل

وحتى نستطيع حساب مجاميع المربعات والجداءات السابقة، قمنا بحساب مجاميع قيم X ومجاميع قيم Y الكلية والهامشية، ووضعناها في الجدول (4–19) السابق, كما قمنا بحساب مجاميع المربعات والجداءات لـ X وY ووضعناها في الجدول (4–20) , وبناء على معطيات هذين الجدولين نجد أن:

جدول (4–20): جدول مساعد لحساب مربعات وجداءات Y و X من عناصر العينات

الرموز	المدرسة الأولى			المدرسة الثانية			المدرسة الثالثة			المجاميع		
	y_{1i}^2	x_{1i}^2	$y_{1i}x_{1i}$	y_{2i}^2	x_{2i}^2	$y_{2i}x_{2i}$	y_{3i}^2	x_{3i}^2	$y_{3i}x_{3i}$	Y^2	X^2	Y*X
1	5184	3025	3960	2500	5776	3800	6721	3844	5084			
2	3600	4900	4200	4356	6400	5280	7744	5100	7920			
3	2025	900	1350	8100	7396	7740	8100	6724	7380			
4	1089	729	891	7596	4900	6020	2500	1600	2000			
5	5100	7921	8010	8281	7225	7735	4900	1764	2940			
6	4624	2500	3400	6400	5329	5840	5625	6400	6000			
7	5929	3600	4620	2500	1600	2000	6400	8100	7200			
8	6400	9604	7840	1600	1225	1400	9025	3600	5700			
9	7225	6241	6715	2209	625	1175	1600	900	1200			
10	4900	6724	5740	8100	3600	5400	2500	1936	2200			
المجموع $\sum_{i=1}^{10}$	49067	46144	46726	51442	44067	46390	55118	72968	47624	155636	133188	140740
متوسط مربعات المجاميع (*)	46240	40960	—	47610	39690	—	51840	38440	—	—	—	—
متوسط جداء المجاميع (**)	—	—	43520	—	—	43470	—	—	44640	—	—	—
$(SSE)_k$	2836	5184	—	3832	4386	—	3278	4528	—	9946	14098	—
$(SPE)_k$	—	—	3206	—	—	2920	—	—	2984	—	—	9110

(*) تم حساب متوسط مربعات المجاميع اعتماداً على الجدول (4–19) ومن العلاقة: $\frac{Y_k^2}{r}$ ثم $\frac{X_k^2}{r}$ مثل: $46240 = \frac{(680)^2}{10} = \frac{Y_1^2}{r}$

(**) كما تم حساب متوسط جداءات المجاميع اعتماداً على الجدول (5–19) ومن العلاقة : $\frac{Y_k X_k}{r}$ مثل: $43520 = \frac{(680)(640)}{10} = \frac{Y_1 X_1}{r}$

$$X_1=\sum_{i=1}^{10}x_{1i}=640 \qquad Y_1=\sum_{i=1}^{10}y_{1i}=680$$
$$X_2=\sum_{i=1}^{10}x_{2i}=630 \qquad Y_2=\sum_{i=1}^{10}y_{2i}=690$$
$$X_3=\sum_{i=1}^{10}x_{3i}=620 \qquad Y_3=\sum_{i=1}^{10}y_{3i}=720$$
$$X=\sum_{k=1}^{3}\sum_{i=1}^{10}x_{ki}=1890 \qquad Y=\sum_{k=1}^{3}\sum_{i=1}^{10}y_{ki}=2090$$

ومن الجدول (4-19) نجد أن: $T_x=\sum\sum x_{ki}^2=133188 \quad ,T_y=\sum\sum y_{ki}^2=155636$

وبالاعتماد على العلاقات من (4-84) إلى(4-88) السابقة وعلى الجدولين (4-19) و(4-20) , قمنا بحساب مجاميع المربعات والجداءات لكل من Y وX كما يلي:

$$(SST)_y=T_y-\frac{Y^2}{gr}=155636-\frac{(2090)^2}{30}=10032.67$$
$$(SSE)_y=\sum_{k=1}^{3}(SSE_k)_y=2836+3832+3278=9946$$
$$(SSA)_y=(SST)_y-(SSE)_y=10032.67-9946=86.67$$
$$(SST)_x=T_x-\frac{X^2}{gr}=133188-\frac{(1890)^2}{30}=14118$$
$$(SSE)_x=\sum_{k=1}^{3}(SSE_k)_x=5184+4386+4528=14098$$
$$(SSA)_x=(SST)_x-(SSE)_x=14118-14098=20$$
$$(SPT)_{xy}=T_{xy}-\frac{X*Y}{gr}=140740-\frac{(2090)(1890)}{30}=74905$$
$$(SPE)_{xy}=\sum_{k=1}^{3}(SSE_k)_{xy}=3206+2920+2984=9110$$
$$(SPA)_{xy}=(SPT)_{xy}-(SPE)_{xy}=74905-9110=65795$$

ولتسهيل الإجراءات العملية نقوم بوضع نتائج هذه الحسابات في جدول مختصر كما يلي:

الجدول (4-21): مربعات وجداءات X وY

مصدر التباين	درجة الحرية	مجموع مربعات X	مجموع مربعات Y	مجموع الجداءات $X*Y$
المجتمعات (المعالجات) SSA	2	20	86.67	65795
الأخطاء (البواقي) SSE	27	14098	9946	9110
الاجمالي SST	29	14118	10032.67	74905

ثم نقوم بحساب المجاميع المعدلة للتابع X من العلاقات التالية :

$$(SST)' = (SST)_x - \frac{(SPT)_{xy}^2}{(SST)_y} = 14118 - \frac{(74905)^2}{10032.67} = 8525.5$$

$$(SSE)' = (SSE)_x - \frac{(SPE)_{xy}^2}{(SSE)_y} = 14098 - \frac{(9110)^2}{9946} = 5753.73$$

$$(SSA)' = (SST)' - (SSE)' = 8525.5 - 5753.73 = 2771.77$$

ثم نقوم بوضع المجاميع المعدلة الأخيرة في جدول تحليل ANCOVA فنحصل على أن:

جدول (4-22) جدول تحليل التغاير ANCOVA حيث أن: $g = 3$ و $r = 10$

مصدر التباين	درجة الحرية	مجاميع المربعات المعدلة	متوسط مجاميع المربعات المعدلة	قيمة F المحسوبة
المجتمعات أو المعالجات SSA	$g - 1 = 2$	2771.77	1385.885	$F = \frac{1385.885}{221.30} = 6.25$
الأخطاء العشوائية SSE'	$g(r-1) - 1 = 26$	5753.73	221.30	ــــــ
Totel الاجمالي	$gr - 2 = 28$	8525.5	ــــــ	ــــــ

ثم نقوم بحساب قيمة مؤشر الاختبار F من العلاقة :

$$F = \frac{MSSA'}{MSSE'} = \frac{1385.885}{221.30} = 6.25$$

ثم نبحث في جداول F عن القيمة الحرجة لمتحول F المقابلة لمستوى الدلالة ($\propto = 0.05$) ولدرجتي الحرية $v_1 = 2$ و $v_2 = 26$ فنجد أن :

$$F_{v_1,v_2}(\propto) = F_{2,26}(0.05) = 3.37$$

وبالمقارنة نجد أن: $(F = 6.25) > 3.37$، لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 ، التي تقول أنه يوجد تأثير للمجتمعات المدروسة على علامات الطلاب في مقرر الرياضيات، وذلك بعد عزل أثر المتحول Y المرتبط مع المتحول X والمتمثل بدرجة الذكاء لهؤلاء الطلاب .

ملاحظة: يستخدم تحليل التباين المشترك ANCOVA لتحقيق هدفين هما:

1- **لزيادة دقة التجربة** وذلك لأنه يعمل على عزل وإبعاد المتحول الإضافي Y المرتبط مع X. وحتى نظهر للقارئ معنى هذا الكلام ننشأ جدول تحليل التباين البسيط ANOVA قبل استبعاد المتحول Y وتعديل المجاميع . فنجد أنه كما يلي:

جدول (4-23): جدول ANOVA قبل التعديل

مصدر التباين	درجة الحرية	مجاميع مربعات الانحرافات		
		المربعات X	المتوسطات XY	F
المجتمعات (SSA)	2	20	10	$F = 0.019$
الأخطاء (SSE)	27	14096	522.07	
الإجمالي (SST)	29	14118		

ومنه نحسب قيمة المؤشر F للمتحول X فنجد أن:

$$F_x = \frac{20/2}{14098/27} = \frac{10}{522.15} = 0.019$$

وبمقارنة هذه القيمة المحسوبة مع القيمة الحرجة $F_{2\,,27}(0.05) = 3.37$ نجد أن $F < 3.37$, لذلك كان يمكن أن نقبل H_0 التي تقول أن $\propto_k = 0$، ونقول أن المجتمعات المدروسة لا تؤثر على علامات الرياضيات. وهكذا نجد أن هذه النتيجة تخالف النتيجة السابقة التي حصلنا عليها بعد عزل تأثير درجة الذكاء Y. وهكذا نجد أن تحليل التباين المشترك ANCOVA يكون مفيداً في تدقيق النتائج وتصويبها .

2- **لتقليل نسبة الخطأ وزيادة الكفاءة النسبية:** ففي مثالنا الحالي نجد أن نسبة الخطأ (من المجموع) قبل التعديل كانت تساوي:

$$P_1 = \frac{SSE_X}{SST_X} = \frac{14098}{14118} = 0.9986$$

أما نسبة الخطأ بعد التعديل (من المجموع المعدل) فتساوي :

$$P_2 = \frac{SSE'}{SST'} = \frac{5753.73}{8525.5} = 0.6749$$

ولتقدير الكفاءة النسبية لتحليل التغاير نقوم بحسابها من العلاقة التالية :

$$RE = \frac{MSSE}{MSSE'} = \frac{SSE/g(r-1)}{SSE'/(g(r-1)-1)} = \frac{1}{1-r^2}$$

$$RE = \frac{14098/27}{5758.73/26} = \frac{523.14}{221.30} = 2.36$$

أي أن تطبيق تحليل التباين المشترك ANCOVA قد أدى إلى زيادة كفاءة التحليل ودقة التجربة بمقدار 2.36 مرة .

ملاحظة: يمكن التوسع في تطبيقات تحليل التباين المشترك لزيادة حساسية التجربة، وذلك باستخدام معيارين وإجراء التحليل باتجاهين للتقليل من أثر المتحول الإضافي Y . وعندها نقوم بنفس الخطوات المذكورة في هذه الفقرة مع إجراء بعض التعديلات اللازمة عليها .

وسنترك ذلك للقارئ للتعمق بها ودراستها في المراجع المختصة .

تمارين الفصل الرابع

1- قامت شعبة الامتحانات بإعداد جداول تتعلق بعلامات بعض الطلبة لعدة مقررات، وكانت النتائج كما يلي (العلامة من مئة) .

المقررات	العينة الأولى	العينة الثانية	العينة الثالثة	العينة الربعة
المقرر الأول	80	65	79	92
المقرر الثاني	60	74	53	71
المقرر الثالث	45	52\4	60	48
المقرر الرابع	52	72	82	42
المقرر الخامس	90	98	76	87

هل يمكننا أن نعد هذه العينات الأربع قد سحبت من مجتمعات ذات أوساط متماثلة إذا كان مستوى الدلالة 5% و 1% .

2- أراد احد المزارعين أن يجرب نوعاً جديداً من الأسمدة وذلك للحكم إذا كان هذا السماد يزيد من إنتاجية الأرض أم لا . ومن أجل هذا قام هذا المزارع بزراعة 12 قطعة من الأرض الزراعية، مساحاتها متساوية، استخدام في /7/ منها السماد الجديد، وفي /5/ منها السماد القديم وقد سجل إنتاجية هذه القطع معبراً عنها بنسب مئوية وفق التالي:

سماد قديم	سماد جديد
85	89
88	90
78	82
70	75
75	70
	92
	86

فهل هناك فرق بين متوسط نسبة الإنتاجية لكل واحدة من هاتين العينتين ضمن مستوى دلالة قدره 5% ؟

3- نريد اختبار إذا ما كان هنالك فرق في المستوى التقني، وذلك مستوى دلالة 5%، لخمس فرق كرة قدم اعتماداً على نتائج اختبار افرادي موفق لـ 6 لاعبين من كل فريق اخترناهم عشوائياً .

ولنفترض أن النتائج التي حصلنا عليها في الاختبار هي كما يلي: (العلامة من 20)

الفريق الأول	الفريق الثاني	الفريق الثالث	الفريق الرابع	الفريق الخامس
15	14	17	19	15
12	15	18	18	17
18	16	16	12	13
14	17	15	17	14
18	15	14	18	18
17	16	18	14	16

4-رغب مدير الإعلان في إحدى دور النشر أن يحدد إذا ما كان هنالك فرق بين متوسط عدد المجلات المباعة لدى مجموعة من مراكز البيع وذلك خلال عدة أسابيع، بالاعتماد على المعلومات التالية:

المركز الرابع	المركز الثالث	المركز الثاني	المركز الأول	الأسبوع
16	36	52	20	1
20	40	44	16	2
32	36	56	32	3
20	40	36	16	4

أ- هيء جدول توزيع التباين .

ب- اختبر فرضية عدم وجود فرق بين متوسط عدد المجلات المباعة لدى مجموعة مراكز البيع ضمن مستوى دلالة قدره 0,01 .

5- لمقارنة المستوى العلمي لثلاث مدارس، عمدنا إلى جمع المعلومات عن علامات عشرة طلبة من كل مدرسة . يمثل الرقم الأول مجموع علامات الطالب في كل المقررات وهي من 260، ويمثل الرقم الثاني علامة الطالب في مقرر الرياضيات وهي من 100، كما يظهر الجدول التالي :

الثالثة		الثانية		الأولى		المدرسة
x	y	x	Y	x	y	
80	210	90	180	80	230	1
70	220	80	170	90	210	2
60	240	70	165	50	130	3
55	160	80	175	60	170	4
45	140	85	160	50	160	5
55	240	65	140	40	150	6
60	230	60	135	30	140	7
40	100	70	135	85	240	8
90	250	95	250	95	220	9
90	210	95	255	70	190	10

المطلوب:

تحديد إذا ما كان هنالك فرق بين مستويات الطلبة، وذلك بالاعتماد على تحليل التباين المشترك ضمن مستوى دلالة قدره 5% .

6-اراد مدير مكتب للضرب على الآلة الكاتبة أن يقارن بين مستويات موظفيه من خلال تجربة بالاعتماد على المربع اللاتيني خلال يوم عمل الذي يعطينا توزيع الموظفين الأربعة وعدد الصفحات المنتجة خلال فترتي عمل 1 و m

l \ m	1	2	3	4
1	2 10	3 8	4 9	1 7
2	4 9	1 7	2 10	3 9
3	1 8	2 6	3 11	4 10
4	3 7	4 9	1 8	2 12

المطلوب:

اختبار هذه الفرضية ضمن مستوى دلالة قدره 1% .

الفصل الخامس

الانحدار الخطي البسيط Simple Linear Regression

5-1: تمهيد:

إن الانحدار الخطي البسيط يدرس العلاقة بين متحولين كميين فقط ، متحول كمي تابع ونرمز له بـ Y، ومتحول كمي مستقل ومؤثر في Y نرمز له بـ X، كالعلاقة بين وزن الطفل Y وعمره X ، أو كالعلاقة بين إنفاق الأسرة Y وعدد أفرادها X، أو كالعلاقة بين الناتج المحلي للدولة Y وعدد السكان فيها X ...الخ .

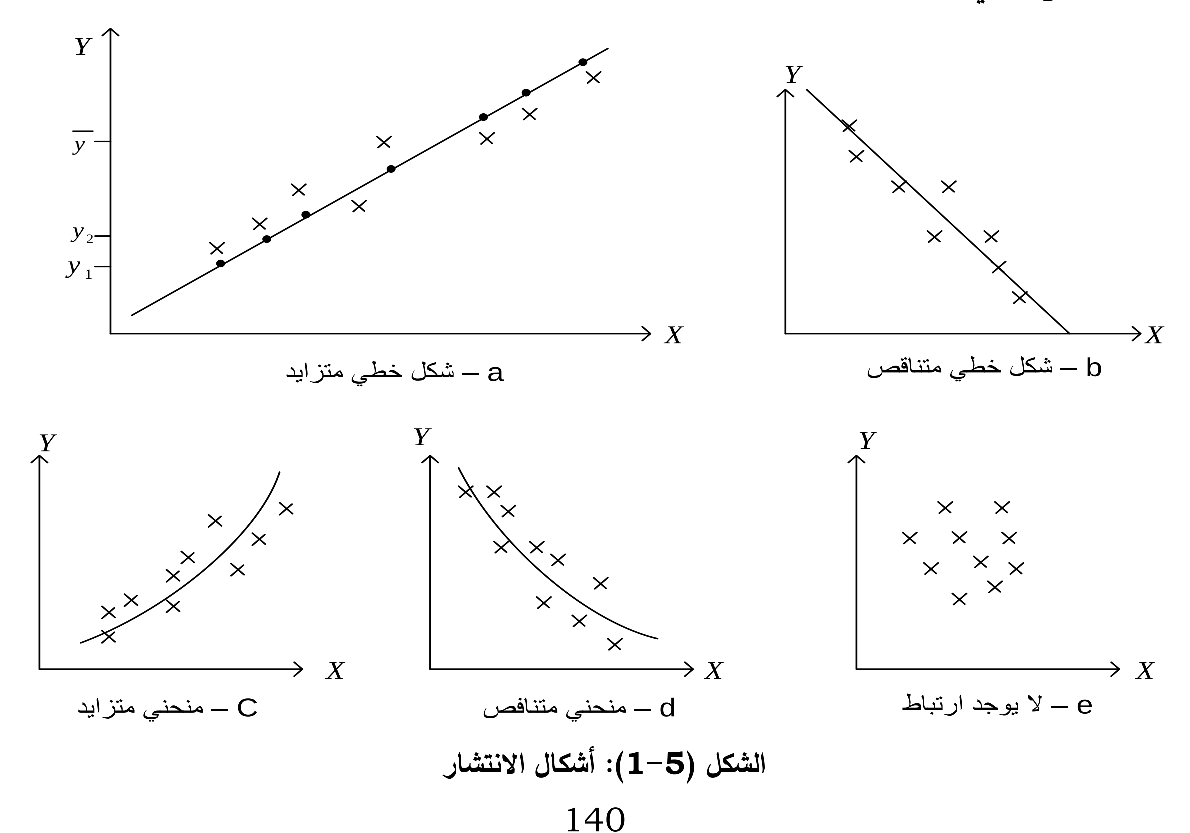
ولدراسة مثل هذه العلاقات في المجتمع المفروض نسحب عينة عشوائية من عناصره بحجم n عنصراً، ونأخذ من كل عنصر فيها القياسات المتقابلة (x_i , y_i) لكل من Y وX. ونضعها في جدول منظم كالتالي :

جدول (5-1): بيانات العينة لـ X وY:

i رقم عنصر العينة	1	2	3	....	i	....	n	المتوسط	الانحراف المعياري
x_i	x_1	x_2	x_3	....	x_i	....	x_n	$\bar{x}$	σ_x
y_i	y_1	y_2	y_3	....	y_i	....	y_n	$\bar{y}$	σ_y

حيث يتم حساب $\bar{x}$ و$\bar{y}$ وσ_x وσ_y من العلاقات المعروفة في الإحصاء .

وعند رسم هذه النقاط المتقابلة (x_i , y_i) على المستوى XoY نحصل على ما يسمى بشكل الانتشار، والذي يأخذ أحد الأشكال التالية :



الشكل (5-1): أشكال الانتشار

ومن هذه الأشكال نلاحظ أنه يوجد لدينا عدة أشكال ممكنة للعلاقة بين Y وX هي:

- الارتباط أو الانحدار الخطي المتزايد والمتناقص كما على الرسمين a وb من الشكل (5−1) .
- الارتباط أو الانحدار المنحني المتزايد والمتناقص كما في الرسمين c وd من الشكل (5−1) .
- عدم وجود ارتباط بين المتحولين Y وX كما في الرسم e من الشكل (5−1) .

5−2: صياغة النموذج:

وسنركز اهتمامنا هنا على الانحدار الخطي المبين في الرسمين a وb من الشكل (5−1) . وسنفترض أن الطبيعة العامة للعلاقة بين Y وX هي علاقة سببية , وإن الصيغة الرياضية لها في المجتمع هي من الشكل الخطي التالي :

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad \text{(النموذج في المجتمع)} \qquad (5-1)$$

حيث أن ε هو حد الخطأ العشوائي . وهو متحول عشوائي مؤلف من حدود مستقلة عن بعضها البعض ومستقلة عن المتحول المستقل X ، وإن توقعها يساوي الصفر $E(\varepsilon) = 0$ ، وإن تباينها ثابت ويساوي σ^2 ، وهي تخضع للتوزيع الطبيعي $N(0\ ,\sigma^2)$.

أما الأمثال β_0 وβ_1 فهي أعداد حقيقية، تعمل على تحديد وضعية المستقيم (5−1) في المستوى XoY وبحيث يكون ميله مساوياً لـ β_1 وقاطعه مع oY مساوياً لـ β_0 . ولكن بما أن الأمثال β_0 وβ_1 مازالت مجهولة فإن وضع المستقيم (5−1) في المستوى يبقى غير محدد .

ولتحديد الوضع المناسب لذلك المستقيم في المستوى نعتمد على بيانات العينة المسحوبة من ذلك المجتمع، ونرسم شكل الانتشار للنقاط الهندسية $(x_i\ , y_i)$، فإذا كان شكلها يوحي لنا باتجاه خط مستقيم نفترض أن العلاقة بين Y وX في العينة هي علاقة خطية أيضاً وتأخذ الشكل التقديري التالي :

$$Y = b_0 + b_1 X + e \qquad \text{(نموذج العينة)} \qquad (5-2)$$

حيث أن e هو حد خطأ التقدير، وهو متحول عشوائي آخر (يختلف عن ε) ومؤلف من حدود مستقلة وخاضعة للتوزيع الطبيعي $N(0\ ,\sigma^2)$. لأن وضعية المستقيم تختلف من عينة لأخرى .

وإذا أخذنا قيمة محددة لـ X مثل x_i، فإننا نحصل على قيمة عددية لـ Y نرمز لها بـ y_i وتساوي :

$$y_i = b_0 + b_1 x_i + e_i \qquad (5-3)$$

وإذا اعتمدنا على معادلة مستقيم العينة $Y = b_0 + b_1 X$ وأخذنا قيمة محددة لـ X مثل x_i فإننا سنحصل مقابلها على قيمة تقديرية لـ Y. نرمز لها بـ $\tilde{y}_i$ ونكتبها كما يلي :

$$\tilde{y}_i = b_0 + b_1 x_i \qquad \text{(قيمة } y_i \text{ التقديرية أو النظرية)} \qquad (5-4)$$

وبتعويض ذلك في (5−3) نجد أن بيانات العينة تعطينا أن:

$$y_i = \tilde{y}_i + e_i \qquad (i : 1\ 2\ 3 \dots n) \qquad (5-5)$$

حيث أن $\tilde{y}_i = b_0 + b_1 x_i$ وهو تقدير y_i . وأن : e_i هو مقدار الخطأ في تقدير y_i . وهو يساوي :

$$e_i = y_i - \tilde{y}_i \qquad (i : 1\ 2\ 3 \dots n) \qquad (5-6)$$

5-3: الافتراضات الموضوعة على النموذج الخطي هي:

1- أن تكون قياسات المتحول المستقل X دقيقة ومحددة (غير عشوائية) .

2- أن يكون المتحول X مؤثراً على Y ويساهم في تفسير تغيراته(أي أن العلاقة بينهما سببية) .

3- أن لا يقل عدد المشاهدات الزوجية عن (3) مشاهدات .

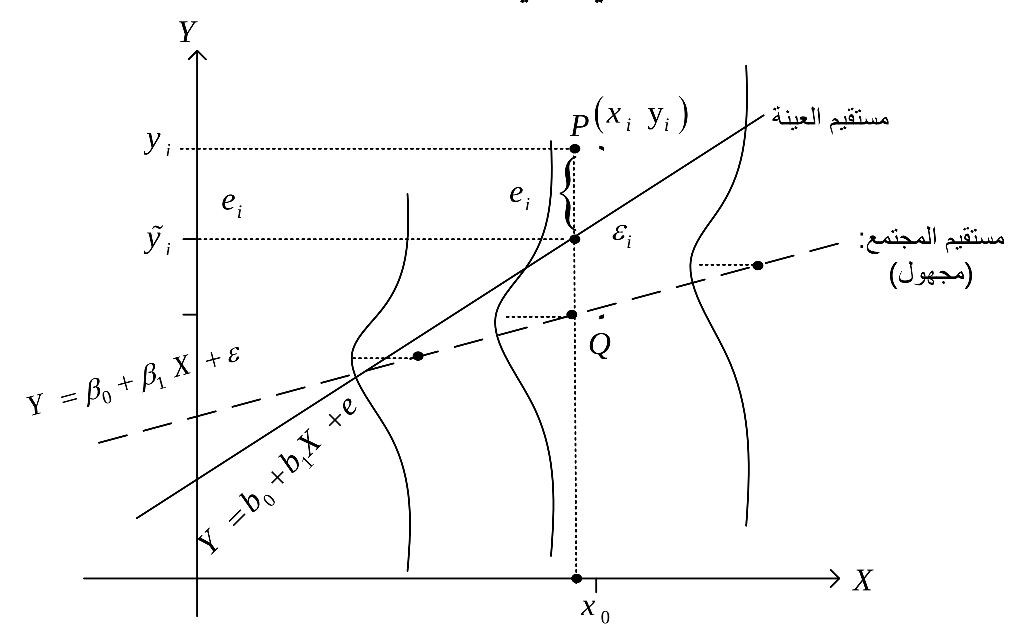
4- أن تكون القيمة المتوقعة لحدود الخطأ العشوائي ε مساوية للصفر $E(\varepsilon) = 0$.

5- أن يكون قيمة تباين حدود الخطأ العشوائي ε ثابتة عند كل قيم X وتساوي σ^2 .

6- أن تكون حدود الخطأ العشوائي ε مستقلة عن بعضها البعض وتخضع للتوزيع الطبيعي $N(0 \; , \sigma^2)$.

7- أن تكون حدود الخطأ العشوائي ε مستقلة عن المتحول المستقل X .

ولتوضيح هذه الأمور نعرضها على الشكل البياني التالي :



الشكل (5-2): مستقيما الانحدار في المجتمع والعينة

5-4: حساب معالم النموذج بطريقة المربعات الصغرى :

حتى يكون مستقيم الانحدار في العينة ممثلاً لشكل الانتشار يجب أن يأخذ أفضل وضعية ممكنة بين نقاط الانتشار، وهذه الوضعية هي التي تجعل مجموع مربعات أخطاء التقدير $\left(\sum_{i=1}^{n} e_i^2\right)$ اصغر ما يمكن، وهذا هو مبدأ طريقة المربعات الصغرى ونكتبه كما يلي :

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 = f(b_0 b_1) \rightarrow min \quad (7-5)$$

ولجعل المقدار $f\left(\tilde{b}_0 \tilde{b}_1\right)$ أصغر ما يمكن نشتقه بالنسبة لـ b_0 ثم بالنسبة لـ b_1 ونضع هذين المشتقين مساويين للصفر فنحصل على ما يلي:

$$\frac{\partial f}{\partial b_0} = 2(y_i - b_0 - b_1 x_i)(-1) = 0$$
$$\frac{\partial f}{\partial b_1} = 2(y_i - b_0 - b_1 x_i)(-x_i) = 0 \qquad (8-5)$$

وبعد إصلاح هذين المشتقين نحصل على المعادلتين العاديتين التاليتين :

$$nb_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$
$$b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (9-5)$$

ثم نحل هاتين المعادلتين الخطيتين اللتين تتضمنان المجهولين b_0 و b_1 , فنحصل على قيمتين محددتين لهما, كما يمكن حسابهما مباشرة من العلاقتين المكافئتين لـ (5−9) التاليتين :

$$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \qquad (10-5)$$

$$b_0 = \bar{y} - b_1 \bar{x} \qquad (11-5)$$

حيث أن $\bar{x}$ و $\bar{y}$ هما متوسطا قيم X و Y على الترتيب .

5−5: حساب معامل الارتباط (البيرسوني) :

لحساب معامل الارتباط (البيروسوني) r_{xy} بين Y و X نستخدم العلاقة المعرفة كما يلي :

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n * \sigma_{\bar{x}} * \sigma_y} = \frac{\sum x_i\, y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}} \qquad (12-5)$$

ومن خواص r_{xy} أن: $-1 \le r_{xy} \le +1$ ، وكلما كانت قيمته قريبة من (1+) كان الارتباط قوياً وطردياً، وكلما كانت قيمته قريبة من (1−) كان الارتباط قوياً وعكسياً . كما أن قيمة r_{xy} لا تتأثر بواحدات القياس المستخدمة في قياس قيم X أو Y .

5−6: حساب القيم النظرية لـ Y .

بعد حصولنا على b_0 و b_1 من المعادلتين (5−9) نعوضهما في معادلة النموذج (5−4) , فنحصل على المعادلة التقديرية التالية:

$$\tilde{Y} = b_0 + b_1 X \qquad (13-5)$$

وبتعويض قيم X فيها نحصل مقابل كل قيمة x_i على قيمة تقديرية نظرية $\tilde{y}_i$ تساوي:

$$\tilde{y}_i = b_0 + b_1 x_i \qquad (14-5)$$

وبعد أن نحصل على القيم النظرية $\tilde{y}_i$, نقوم بحساب الأخطاء $e_i = (y_i - \tilde{y})$ ثم نربعها ونضعها في جدول مناسب كما يلي :

جدول (5-2): جدول مساعد لإجراء الحسابات اللازمة

الانحراف المعياري	المتوسط	n	...	I	...	3	2	1	رقم العنصر i
σ_x	$\bar{x}$	x_n	...	x_i	...	x_3	x_2	x_1	x_i
σ_y	$\bar{y}$	y_n	...	y_i	...	y_3	y_2	y_1	y_i
——	$\bar{\tilde{y}} = \bar{y}$	$\tilde{y}_n$	...	$\tilde{y}_i$	...	$\tilde{y}_3$	$\tilde{y}_2$	$\tilde{y}_1$	$\tilde{y}_i$
——	0	$(y_n - \tilde{y}_n)$	...	$(y_i - \tilde{y}_i)$	...	$(y_3 - \tilde{y}_3)$	$(y_2 - \tilde{y}_2)$	$(y_1 - \tilde{y}_1)$	$e_i = (y_i - \tilde{y}_i)$
——	——	$(y_n - \tilde{y}_n)^2$		$(y_i - \tilde{y}_i)^2$		$(y_3 - \tilde{y}_3)^2$		$(y_1 - \tilde{y}_1)^2$	$(y_i - \tilde{y}_i)^2$

ومن الجدول (5-2) نحسب الخطأ المعياري للتقديرات من العلاقة التالية :

$$S_{y/x} = \sqrt{\frac{\sum(y_i - \tilde{y}_i)^2}{n-2}} = \sqrt{\frac{\sum e_i^2}{n-2}} \qquad \mathbf{(15-5)}$$

ولتقدير جودة التمثيل نقوم بحساب معامل التحديد R^2 العادي من العلاقة :

$$R^2 = 1 - \frac{\sum e_i^2}{\sum(y_i - \bar{y})^2} \qquad (16-5)$$

كما يمكن حساب معامل التحديد المعدل من العلاقة :

$$Ad_jR^2 = 1 - \frac{S_{y/x}^2}{\sigma_y^2} \qquad (17-5)$$

وكلما كانت قيمة R^2 قريبة من الواحد كان التقدير جيداً .

كما يمكننا اختبار معنوية التقديرات b_0 و b_1 و r_{xy} و R^2 حسب القواعد المعروفة في الاقتصاد القياسي. وسنتعرض لهما في الانحدار المتعدد .

5-7: اختبار جودة التمثيل وتحليل التباين:

لاختبار صلاحية النموذج نستخدم تحليل التباين ANOVA لدراسة تغيرات Y بدلالة X . ولذلك نقوم بتحليل مجموع مربعات انحرافات Y عن متوسطها الحسابي $\bar{y}$ كما يلي:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \tilde{y}_i + \tilde{y}_i - \bar{y})^2$$
$$\sum(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 + \sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2 + 2(0) \qquad \mathbf{(18-5)}$$

ومن جهة أخرى نلاحظ أن الحد الأول يساوي $\sum_{i=1}^{n} e_i^2$، أما الحد الثاني فنعالجه بتعويض $\tilde{y}_i$ و $\bar{y}$ كما يلي :

$$\sum_{i=1}^{n}(\tilde{y}_i - \bar{y}) = \sum_{i=1}^{n}(\, b_0 + b_1x_i - \, b_0 - b_1\bar{x})^2 = b_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2$$

وبذلك نجد (بعد تبديل موقعي الحدين) أن (5-18) تأخذ الشكل التالي:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = b^2\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n} e_i^2 \qquad \mathbf{(19-5)}$$

وباستخدام الرموز المعروفة لهذه المجاميع نجد أن:

$$SST = SSX + SSE \qquad (20-5)$$

وإن درجات الحرية المقابلة لها هي كما يلي:

$$n-1 = 1 + (n-2) \qquad (21-5)$$

ثم نقوم بوضع فرضتي الاختبار كما يلي:

$H_0 : b_1 = 0$ فرضية العدم (النموذج غير صالح)

$H_1 : b_1 \neq 0$ الفرضية البديلة　　(22 − 5)

وبعد إجراء الحسابات اللازمة نضع النتائج في جدول ANOVA كما يلي:

جدول (5-3): جدول ANOVA **لاختبار علاقة** Y بـ X:

مصدر التباين	درجة الحرية v	مجموع المربعات	متوسط المربعات	قيمة F
المتحول X	1	$SSX = \sum^{n} (\tilde{y}_i - \bar{y})^2 = b_1^2 \sum (x_i - \bar{x})^2$	$MSX = \frac{SSX}{1} = b_1^2 \sum (x_i - \bar{x})^2$	$F = \frac{MSX}{MSE}$
الأخطاء أو البواقي	$n-2$	$SSE = \sum^{n} (y_i - \tilde{y}_i)^2 = \sum^{n} e_i^2$	$MSE = \frac{SSE}{n-2}$	
الاجمالي	$n-1$	$SST = \sum (y_i - \bar{y})^2$		

واعتماداً على بيانات الجدول (5-3) نقوم باختبار صلاحية التمثيل باستخدام المؤشر F المعرف بالعلاقة :

$$F = \frac{MSX}{MSE} = \frac{b_1^2 \sum (x_i - \bar{x})^2}{\sum e_i^2 / (n-2)} \qquad (23-5)$$

وهو متحول عشوائي يخضع لتوزيع F بدرجتي حرية $v_1 = 1$ و$v_2 = n-2$، لذلك نقوم بمقارنة قيمة F المحسوبة من (5-23) مع القيمة الحرجة لمتحول $F_{v_1,v_2}(\propto)$ عند مستوى الدلالة $\propto$ ودرجتي الحرية v_1 وv_2، ونتخذ القرار كما يلي :

إذا كانت $F \leq F_{1\ ,n-2}(\propto)$ نقبل فرضية العدم H_0 . التي تقول أن أمثال X في النموذج معدومة $b_1 = 0$، ونستنتج أن النموذج لا يمثل العلاقة المدروسة بين Y وX .

أما إذا كانت $F > F_{1\ ,n-2}(\propto)$ نرفض فرضية العدم H_0 ونقبل الفرضية التي تقول أن $b_1 \neq 0$ ونقول أن النموذج يمثل العلاقة بين Y وX باحتمال $(1 - \propto)$.

5-8: الاستدلال من تابع الانحدار المقدر:

وهو يستخدم لمعالجة مسائل التنبؤ بقيمة Y المقابلة لقيمة معينة لـ X مثل x_0 . علماً بأن عملية التنبؤ تعني إيجاد تقدير لقيمة Y المجهولة ثم إيجاد مجال الثقة الذي يحتوي هذه القيمة باحتمال ثقة$(1-\propto)$.

لذلك نفترض إننا اخترنا قيمة لـ X مثل x_0 . ونريد أن نتنبأ بقيمة Y المقابلة لها، وإذا ذهبنا إلى الشكل القادم (5-3) نلاحظ أنه مقابل أية قيمة x_0 يوجد لدينا (3) قيم لـ Y هي:

1- القيمة الفعلية أو الحقيقية لـ Y ونرمز لها بـ y_0 ، وهي قيمة مجهولة وتقابل النقطة $p(x_0, y_0)$ على الشكل (5-3)، وهي تساوي حسب النموذج العام ما يلي:

$$Y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0 \qquad (24-5)$$

2-القيمة المتوقعة لـ y_0 : لأن y_0 هي عبارة عن متحول عشوائي يخضع في المجتمع لتوزيع احتمالي معين، ويمكن أن يأخذ أية قيمة ممكنة له (حسب العينات المختلفة). وتتمركز هذه القيم حول توقعها الرياضي $E(y_0)$، وهو يقع على مستقيم الانحدار في المجتمع ويقابل القيمة x_0، وإن هذا التوقع مجهول أيضاً . ولكنه يساوي حسب النموذج العام ما يلي :

$$E(y_0) = E(\beta_0 + \beta_1 x_0 + \varepsilon_0) = \beta_0 + \beta_1 x_0 = \left(\text{مستقيم المجتمع}\right) \qquad (25-5)$$

وهي نقطة تقع على مستقيم المجتمع (المجهول أيضاً) وتقابل القيمة x_0 ، ولقد رمزنا لها على الشكل (5-3) بالرمز Q ، وهي تمثل مركز التوزيع الاحتمالي الشرطي لـ y_0 المقابلة لـ x_0 .

3- القيمة التقديرية لـ y_0: ونرمز لها بـ $\widetilde{y_0}$ وهي قيمة تقديرية لـ y_0 وتحسب من معادلة مستقيم الانحدار للعينة بعد حسابه أصولاً وتساوي ما يلي:

$$\widetilde{y}_0 = b_0 + b_1 x_0 = \left(\text{مستقيم العينة}\right) \qquad (26-5)$$

وهي قيمة معلومة وتقع على مستقيم العينة , وهي تعتبر تقديراً غير متحيز للقيمة الفعلية y_0 (حسب خواص تقديرات المربعات الصغرى) .

وإذا أخذنا توقع هذه القيمة التقديرية (حسب العينات المختلفة) نجد أن:

$$E(\widetilde{y}_0) = E(b_0 + b_1 x_0) = E(b_0) + E(b_1)x_0 = \beta_0 + \beta_1 x_0 \qquad (27-5)$$

وذلك لأن $E(b_0) = \beta_0$ و $E(b_1) = \beta_1$ (حسب تقديرات المربعات الصغرى) وبمقارنة العلاقتين (5-25) و(5-27) نحصل على العلاقة التالية:

$$E(\widetilde{y}_0) = \beta_0 + \beta_1 x_0 = E(Y_0) \qquad (28-5)$$

أي أن التقدير $\widetilde{Y}_0$ هو تقدير غير متحيز للتوقع $E(y_0)$ أيضاً، لأن توقع $\widetilde{y_0}$ يساوي $E(y_0)$.

وهكذا نستنتج مما سبق إنه مقابل كل قيمة x_0 يكون لدينا قيمتان مجهولتان هما: القيمة الفعلية y_0، والقيمة المتوقعة لها $E(y_0)$, ويكون لدينا قيمة معلومة هي القيمة التقديرية $\widetilde{y_0}$ المحسوبة من معادلة الانحدار(5-26) من بيانات العينة المدروسة, وسنستفيد من هذه القيمة المعلومة $\widetilde{y_0}$ في عمليات التنبؤ بقيمة y_0 الفعلية وبقيمة القيمة المتوقعة لها $E(y_0)$. وذلك لأنه قد وجدنا من العلاقتين (5-26) و(5-28) أن $\widetilde{y_0}$ هي

تقدير غير متحيز لكل من y_0 الفعلية وللقيمة المتوقعة لها $E(y_0)$. لذلك سنعتمد عليها في عملية التنبؤ بقيمة y_0وبقيمة $E(y_0)$ كما يلي:

أ- التنبؤ بالقيمة المتوقعة لـ y_0 وهي $E(y_0)$ المقابلة لـ x_0 :

نقوم أولاً بحساب القيمة التقديرية لها $\widetilde{y_0}$ من النموذج المقدر (5-4) فنجد أن:

$$\widetilde{y_0} = b_0 + b_1 x_0 \qquad (29-5)$$

ثم نقوم بحساب الانحراف المعياري لها من العلاقة التالية (انظر البرهان عند العشعوش، والعربيد. ص113) .

$$Sd(\widetilde{y_0}) = S_{y/x} * \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} \qquad (30-5)$$

حيث أن: $S_{y/x}$ هو الخطأ المعياري المعرف بالعلاقة (5-15), ثم نقوم بإنشاء مجال الثقة ذي الاحتمال $(1 - \propto)$ لـ $E(y_0)$ من العلاقة التالية :

$$\widetilde{y_0} - t_{n-2}\left(\frac{\propto}{2}\right) * Sd(\widetilde{y_0}) \le E(y_0) \le \widetilde{y_0} + t_{n-2}\left(\frac{\propto}{2}\right) * Sd(\widetilde{y_0}) \qquad (31-5)$$

وهو مجال يحتوي على القيمة المتوقعة $E(y_0)$ باحتمال $(1 - \propto)$. ولكن هذا المجال يزداد اتساعاً كلما ابتعدنا عن المتوسط $\bar{x}$، وذلك بسبب الحد $(x_0 - \bar{x})^2$.

ب- التنبؤ بالقيمة الفعلية y_0المقابلة لـ x_0 :

نقوم أولاً بإيجاد تقدير القيمة الفعلية y_0المقابلة لـ x_0 من النموذج المقدر (1-4) فنجد أيضاً أن:

$$\widetilde{y_0} = b_0 + b_1 x_0 \qquad (32-5)$$

ثم نقوم بحساب الانحراف المعياري للتقدير $\widetilde{y_0}$ الذي يتضمن في هذه الحالة مصدرين للخطأ هما:

- خطأ تقدير القيمة المتوقعة $E(y_0)$ بواسطة $\widetilde{y_0}$.
- خطأ القياس الفردي لـ y_0 والذي تباينه σ_e^2 والذي يقدر بتباين الأخطاء $S_{y/x}^2$.

وبالتالي نستنتج أن مجال الثقة للقيمة الفعلية y_0 يجب أن يكون أوسع من مجال الثقة للقيمة المتوقعة $E(y_0)$. ولهذا فإن الانحراف المعياري للتقدير $\widetilde{y_0}$ – في هذه الحالة- يحسب من العلاقة التالية: (انظر البرهان عند العشعوش، والعربيد ص116) .

$$SSd(\widetilde{y_0}) = S_{y/x} * \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_0 - \bar{x})^2}} \qquad (33-5)$$

حيث أن: $S_{y/x}$ هو الخطأ المعياري المعرف بالعلاقة (1-15)

ثم نقوم بإنشاء مجال الثقة للقيمة الفعلية y_0 ذي الاحتمال $(1 - \propto)$ من العلاقة التالية:

$$\widetilde{y_0} - t_{n-2}\left(\frac{\propto}{2}\right) * SSd(\widetilde{y_0}) \le y_0 \le \widetilde{y_0} + t_{n-2}\left(\frac{\propto}{2}\right) * SSd(\widetilde{y_0}) \qquad (34-5)$$

وهو مجال يحتوي على القيمة الفعلية y_0 باحتمال $(1 - \propto)$. وإنه يزداد اتساعاً كلما ابتعدنا عن المتوسط $\bar{x}$ بسبب الحد $(x_0 - \bar{x})^2$.

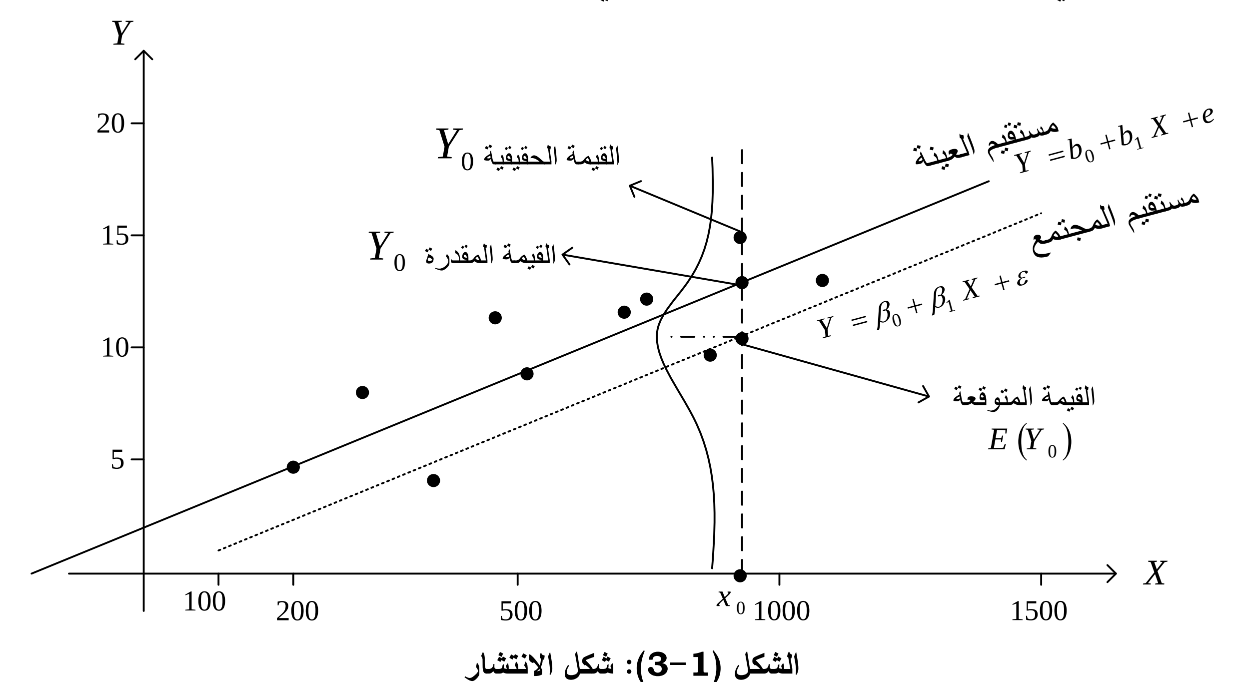
ملاحظة: عند إجراء عمليات التنبؤ الخارجي يجب أن لا نبتعد كثيراً عن حدود نطاق بيانات العينة . لأن ذلك يزيد الخطأ المرتكب ويقلل من دقة النتائج .

مثال (5-1): أدرس العلاقة بين مقدار دخل الأسرة X (بالدولار) وكمية استهلاكها من اللحم Y (كغ/شهريا) ثم تنبأ بكمية الاستهلاك عندما $X = 800$ و$X = 1200$ ، وذلك اعتماداً على بيانات عينة مؤلفة من $(n = 10)$ أسر مختلفة والمبينة في الجدول التالي :

جدول (5-4): بيانات العينة (المصدر فرضي)

رقم الاسرة	1	2	3	4	5	6	7	8	9	10	المتوسط	الانحراف المعياري
دخل الاسرة X$	210	290	350	480	490	730	780	850	920	1010	611	282.51
كمية الاستهلاك Y kg	5	7	6	11	8	11	12	8	15	12	9.5	3.171

الحل: قبل كل شيء نرسم شكل الانتشار فنجد أنه كما يلي:



الشكل (1-3): شكل الانتشار

ومن الجدول (5-4) نجد أن $\bar{x} = 611$ و $\bar{y} = 9.5$ و $\sigma_x = 282.51$ و $\sigma_y = 3.171$

ونفترض أن العلاقة بين Y وX في العينة هي علاقة خطية من الشكل :

$$Y = b_0 + b_1X + e$$

وحتى نستطيع حساب المعلمتين b_0 وb_1 من المعادلتين (5-9) أو من المعادلتين (5-10) و(5-11) ، علينا أن نقوم بإعداد الحسابات الواردة في الجدول التالي :

جدول (5-5): الحسابات المساعدة

رقم الأسرة i	1	2	3	4	5	6	7	8	9	10	المجموع
x_i^2	44100	48100	122500	250400	240100	532900	608400	7225000	896400	1020100	4451500
y_i^2	25	49	36	121	64	121	144	64	225	144	993
$x_i y_i$	1050	2050	2100	5280	5920	8030	9360	6800	13800	12120	64490
القيم النظرية $\tilde{y}_i = b_0 + b_1 x_i$	5.91	6.62	7.16	8.33	8.41	10.57	11.02	11.64	12.27	13.08	95
$e_i^2 = (y_i - \tilde{y}_i)^2$	0.8281	0.1444	1.3456	7.1209	0.1681	0.1849	0.9604	13.2496	6.4529	1.1664	32.6432

نعوض نتائج هذه الحسابات في المعادلتين (5−10) و(5−11)، فنجد أن :

$$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{64490 - 10(611)(9.5)}{4451500 - 10(611)^2} = \frac{6445}{718290} = 0.00897$$

$$b_0 = \bar{y} - b_1\bar{x} = 9.5 - (0.00897)(611) = 4.019$$

وبذلك نحصل على أن معادلة العلاقة الانحدارية تأخذ الشكل التالي :

$$\tilde{y}_i = 4.019 + (0.00897)x_i$$

ومن هذه العلاقة حسبنا القيم النظرية $\tilde{y}_i$ ووضعناها في السطر الخامس من الجدول (5−1) السابق, ثم قمنا بحساب مربعات الفروقات $(y_i - \tilde{y}_i)^2$ ووضعناها في السطر الأخير فوجدنا أن :

$$\sum (y_i - \tilde{y}_i)^2 = \sum e_i^2 = 32.6432 \quad \Longrightarrow \quad S_{y/x}^2 = \frac{32.6432}{8} = 4.0804$$

ومنها نحسب معاملي التحديد العادي والمعدل بعد حساب المقام لهما كما يلي:

$$\sum_{i=1}^{10} (y_i - \bar{y})^2 = \sum_{i}^{10} y_i^2 - n\bar{y} = 90.5$$

فيكون لدينا:

$$\sigma^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{90.5}{10} = 9.05$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{32.6432}{90.5} = 0.639 \approx 0.64$$

ثم نحسب معامل التحديد المعدل من العلاقة:

$$Ad_j R^2 = 1 - \frac{S_{y/x}^2}{\sigma_y^2} = 1 - \frac{4.0804}{9.05} = 0.594$$

ومنهما نستنتج أن جودة التمثيل جيدة (لأن $R^2 > 0.50$). كما يمكننا حساب معامل الارتباط r_{yx} من العلاقة :

$$r_{yx}=\frac{\sum x_i y_i-n\bar{x}\bar{y}}{(n-1)*\sigma_x\sigma_y}=\frac{64490-10(611)(9.5)}{9*(282.51)(9.17)}\approx 0.80$$

وهنا نلاحظ أن مربع معامل الارتباط يساوي معامل التحديد أي أن $(r^2=R^2)$.

وأخيراً نقوم بإجراء تحليل التباين فنجد أن المجاميع المطلوبة تساوي:

$$SST=\sum_{i=1}^{10}(y_i-\bar{y})^2=(n-1)\sigma_y^2=9(3.171)^2=90.5$$

$$SSE=\sum(y_i-\tilde{y}_i)^2=\sum e_i^2=32.6432$$

$$SSX=SST-SSE=90.5-32.6432=57.86$$

وبذلك نجد أن جدول تحليل التباين ANOVA لهذا النموذج يأخذ الشكل التالي :

جدول (5-6): تحليل التباين ANOVA

مصدر التباين	الرمز	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F
المتحول X	SSX	57.85	1	57.80	$F=\frac{57.86}{4.0804}=14.18$
الأخطاء (البواقي)	SSE	32.64	8	4.0804	
الاجمالي	SST	90.50	9	–	

ثم نقوم بإيجاد قيمة $F(\propto)$ الحرجة المقابلة لمستوى الدلالة $\propto=0.05$ ولدرجتي الحرية $v_1=1$ و$v_2=8$ فنجد أنها تساوي :

$$F_{v_1,v_2}(\propto)=F_{1,8}(0.05)=5.32$$

وبمقارنة قيمة F المحسوبة مع قيمتها الحرجة $F(\propto)$ نجد أن: $F=14.18>5.32$، لذلك نرفض فرضية العدم $(H_0: b_1=0)$ ونقبل $(H_1: b_1\neq 0)$ ونعتبر النموذج صالحاً لتمثيل العلاقة بين Y وX.

ولإجراء عملية التنبؤ للقيمة المتوقعة لـ Y المقابلة لقيمة X مثل: $x_0=800$، نعوضها في معادلة النموذج فنحصل على قيمة Y المقدرة :

$$\tilde{y}_{800}=4.019+(0.00897)(800)=11.195 \quad \text{(كغ / شهريا)}$$

ولكن بما أن هذه القيمة المتوقعة لـ y_{800} مجهولة. فإننا نقوم بإنشاء مجال الثقة لها ذي الاحتمال (0.95) من العلاقة (5-31) فنجد أن القيمة المتوقعة لـ y_{800} تتراوح في المجال التالي :

$$\tilde{y}_{800}-t_{n-2}\left(\frac{\propto}{2}\right)*Sd(\tilde{y}_{800})\leq E(y_{800})\leq \tilde{y}_{800}+t_{n-2}\left(\frac{\propto}{2}\right)*Sd(\tilde{y}_{800})$$

ولذلك نحسب $Sd(\tilde{y}_{800})$ من العلاقة (5-25) التالية:

$$Sd(\tilde{y}_{800})=S_{y/x}*\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}=\sqrt{4.0804}\sqrt{\frac{1}{10}+\frac{(800-611)^2}{718290}}=0.7816$$

وذلك لأن:

$$\sum_{i=1}^{10}(x_i-\bar{x})^2=\sum x_i^2-n(\bar{x})^2=4451500-10(611)^2=718290$$

ومن جداول(ستودينت) نجد أن القيمة الحرجة $t_{n-2}\left(\frac{\propto}{2}\right)$ تساوي: $t_{n-2}\left(\frac{\propto}{2}\right)=t_8(0.025)=2.306$

, وبالتعويض في مجال الثقة نجد أن:

$$11.195-(2.306)(0.7816)\leq E(y_{800})\leq 11.195+(2.306)(0.7816)$$
$$9.39\leq \mathrm{E}(y_{800})\leq 12.997\approx 13$$

ولإجراء عملية تنبؤ للقيمة الفعلية لـ (y_{800}) المقابلة لـ$x_0=800$، نعوضها في معادلة النموذج فنجد أن:

$$\tilde{y}_{800}=4.019+(0.00897)(800)=11.195 \quad \text{(كغ / شهريا)}$$

ولإنشاء مجال الثقة للقيمة الحقيقية لـ y_{800} نطبق العلاقة (5–34) فنجد أن:

$$11.195-(2.306)\sqrt{4.0804}\sqrt{1+\frac{1}{10}+\frac{(800-611)^2}{718290}}\leq y_{800}\leq 11.195+(2.306)\sqrt{4.0804}\sqrt{1+\frac{1}{10}+\frac{(800-611)^2}{718290}}$$

$$11.195-4.995\leq y_{800}\leq 11.195+4.995$$
$$6.200\leq y_{800}\leq 16.195$$

وهو مجال يتضمن القيمة الفعلية y_{800} المقابلة لـ $x=800$ باحتمال قدره $1-\propto=0.95$.

وهنا نلاحظ أن مجال الثقة لـ y_{800} الفعلية أوسع من مجال الثقة للقيمة المتوقعة لها $E(y_{800})$. لأن القيمة الفعلية Y_0 أكثر تشتتاً من القيمة المتوقعة $E(y_{800})$. وللتنبؤ بقيمة الاستهلاك Y عندما يكون $X=1200$, ندعو القارئ أن يقوم بها على سبيل التمرين، وعليه أن يتبع نفس الخطوات السابقة .

5–9: حساب تقدير معالم نموذج الانحدار تحت قيود خطية عليها :

لنفترض أن نموذج الانحدار الخطي العام (في المجتمع) كان على الشكل التالي:

$$Y=\beta_0+\beta_1X+\varepsilon \qquad (35-5)$$

وإذا سحبنا عينة عشوائية من عناصر المجتمع وأخذنا قياسات X منها وقياسات Y المقابلة لها، فإننا سنحصل من كل عنصر i منها على قيمة x_i وعلى قيمة y_i . وإذا عوضنا هذه القيم في معادلة النموذج (5–35) نحصل على n معادلة خطية كما يلي:

$$\begin{aligned}y_1&=\beta_0+\beta_1x_1+\varepsilon_1\\ y_2&=\beta_0+\beta_1x_2+\varepsilon_2\\ &\dots\dots\dots\dots\dots\\ y_n&=\beta_0+\beta_1x_i+\varepsilon_n\end{aligned} \qquad (36-5)$$

حيث : ε_i هو حد الخطأ العشوائي المقابلة للعنصر i, ولتسهيل المعالجات الرياضية نكتب هذه المعادلات بدلالة الأشعة والمصفوفات على الشكل التالي :

$$\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\begin{bmatrix}1&x_1\\1&x_2\\\vdots&\vdots\\1&x_n\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix} \qquad (37-5)$$

وبذلك يمكننا كتابة النموذج (5–35) مصفوفياً على الشكل التالي:

$$Y_{n*1} = X_{n*2} * \beta_{2*1} + \varepsilon_{n*1} \qquad (38-5)$$

حيث أن: X هي المصفوفة الموسعة لـ X وتتضمن n سطراً وعمودين. وإن عمودها الأول يتألف من واحدات حتى يتوافق مع المعلمة الثابتة β_0 : أي أن :

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad (39-5)$$

وعندما يتم تقدير المعلمتين β_0 و β_1 من بيانات العينة من إحدى العلاقتين (5−9) أو (5−10), فإننا سنحصل على تقديرهما b_0 و b_1, وعندها فإن صيغة النموذج تأخذ الشكل التالي :

$$Y = b_0 + b_1 X + e = X * b + e \qquad (40-5)$$

حيث أن: $b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$، وأن e هو شعاع الأخطاء في العينة وهو يختلف عن ε، ولكنه يخضع للتوزيع الطبيعي $N(0\,, \sigma^2)$.

والآن نفترض أن هناك مجموعة من القيود أو الشروط مفروضة على معالم النموذج المجهولةβ_0 و β_1 مثل:

$$\beta_0 = 10 \quad , \quad \beta_0 + \beta_1 = 5 \quad , \quad \beta_0 - \beta_1 = 0$$

وهنا نلاحظ أنه يمكننا كتابة هذه القيود مصفوفياً على الترتيب على الشكل التالي :

$$[1\ \ ,0] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 10 \qquad [1\ \ ,1] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 5 \qquad [1\ \ ,-1] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 0 \qquad (41-5)$$

وبصورة عامة يمكننا كتابة هذه القيود على شكل معادلة مصفوفية واحدة على الشكل التالي :

$$R_{g*2} * \beta_{2*1} = C_{g*1} \Rightarrow \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 5 \\ 0 \end{bmatrix} \qquad (42-5)$$

حيث أن R: هي مصفوفة معاملات القيود في معادلاتها الخطية وهي مؤلفة من g سطراً وعمودين . حيث g عدد القيود المفروضة. وأن C: هو عمود الثوابت التي في الطرف الثاني لمعادلات القيود, أي أنه يمكننا كتابة القيود المفروضة على معالم النموذج β بشكل منفرد أو بشكل جماعي ومشترك كما في العلاقتين (5−41) و(5−42) على الترتيب . وهكذا نجد أن المصفوفة R والعمود C يساويان (في هذه الحالة) ما يلي:

$$R = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad C = \begin{bmatrix} 10 \\ 5 \\ 0 \end{bmatrix}$$

كما يمكننا كتابة أي شرط من القيود السابقة بمفرده على الشكل التالي :

$$[1\ \ ,0] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 10 \quad : \quad R_1\, \beta = 10$$

$$[1\ \ ,1] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 5 \quad : \quad R_2\, \beta = 5$$

$$[1\ \ ,-1] * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = 0 \quad : \quad R_3\, \beta = 0$$

والآن علينا أن نقوم بتقدير المعالم β_1 و β_0 تحت القيود R المذكورة في (5−42)، علماً بأن النموذج المقدر بدون قيود يأخذ الشكل التالي :

$$Y = b_0 + b_1 X + e \qquad (43-5)$$

حيث يتم حساب التقديرين b_1 و b_0 بطريقة المربعات الصغرى من بيانات العينة قبل الأخذ بعين الاعتبار الشروط المفروضة .

ولتقدير المعلمتين β_1 و β_0 تحت القيود R، نرمز لتقديرهما المقيد بالرمزين: $\tilde{\beta}_{1R}$ و $\tilde{\beta}_{0R}$, وعندها يمكننا كتابة نموذج الانحدار المقيد والمقدر من العينة على الشكل التالي :

$$Y_R = \tilde{\beta}_{0R} + \tilde{\beta}_{1R} X + e_R = X * \tilde{\beta}_R + e_R \qquad (44-5)$$

حيث أن: $\tilde{\beta}_R$ هو عمود التقديرات المقيدة بالقيود المفروضة .

وأن: e_R هو عمود الأخطاء المقيدة بالقيود المفروضة .

ويمكن حساب الشعاع المقيد $\tilde{\beta}_R$ بعد حساب الشعاع غير المقيد b من بيانات العينة , وذلك باستخدام العلاقة التالية [انظر البرهان عند عناني ص 703] .

$$\tilde{\beta}_R = b - (X'X)^{-1} * R'[R(X'X)^{-1}R'] \, [R \;\; b - C] \qquad (45-5)$$

ومنها نحصل على عمود التقديرات المقيدة $\tilde{\beta}_R$ بالاعتماد على طريقة المربعات الصغرى وبدلالة التقديرات غير المقيدة b .

علماً بأن التقديرات المقيدة تكون تقديرات غير متحيزة للمعالم β المجهولة، إذا كانت القيود $(R\beta = C)$ صحيحة، وتكون متحيزة إذا كانت تلك القيود غير صحيحة . كما أنها تكون أكثر كفاءة من التقديرات b إذا كانت القيود صحيحة, لأن مصفوفة التباينات المشتركة للشعاع $\tilde{\beta}_R$ في هذه الحالة تكون **أقل أو تساوي من مصفوفة التباينات المشتركة للتقديرات** b، أي أن:

$$COV(\tilde{\beta}_R) \le COV(b) \qquad \text{(إذا كانت القيود صحيحة)} \qquad (46-5)$$

وهذا يعني إن فرض بعض القيود الخطية الصحيحة على معالم النموذج β يجعل مقدرات تلك المعالم المقيدة $\tilde{\beta}_R$ أكثر كفاءة لأنها ستكون أقل تبايناً .

ولاختبار معنوية القيود الخطية المفروضة في (5–42) نضع الفرضيتين كما يلي:

$$H_0 : R\beta = C \qquad \text{(القيود صحيحة)}$$
$$H_1 : R\beta \ne C \qquad \text{(القيود غير صحيحة)} \qquad \mathbf{(47-5)}$$

ثم نحسب مؤشر الاختبار F من العلاقة التالية :

$$F = \frac{\frac{[\boldsymbol{e'_R} * \boldsymbol{e_R} - \boldsymbol{e'} * \boldsymbol{e}]}{g}}{\frac{\boldsymbol{e'} * \boldsymbol{e}}{n-k}} \sim F_{g,n-k}(x) \qquad \mathbf{(48-5)}$$

وهو يخضع للتوزيع $F(x)$ وبدرجتي حرية: $v_1 = g$ و $v_2 = n - k$

حيث أن: k هو عدد المعالم β في النموذج .

وأن: $\boldsymbol{e'_R} * \boldsymbol{e_R}$ هو يمثل مجموع مربعات الأخطاء في النموذج المقيد e_R .

وأن: $\boldsymbol{e'} * \boldsymbol{e}$ هو مجموع مربعات الأخطاء في النموذج غير المقيد e .

وأن g عدد الشروط المفروضة على المعالم β .

علماً بأنه يمكن استبدال البسط في العلاقة (5–48) بما يساويه كما يلي:

$$(e_R' * e_R - e' * e) = [Rb - C]'[R(X'X)^{-1}R']^{-1} * [Rb - C] \qquad (49-5)$$

وعندها تأخذ العلاقة (5–48) لمؤشر الاختبار F الشكل التالي :

$$F = \frac{\frac{[Rb - C]'[R(X'X)^{-1}R']^{-1} * [Rb - C]}{g}}{\frac{e' * e}{n-k}} \qquad (50-5)$$

وهي علاقة تعطينا قيمة F بدلالة بيانات العينة X والتقديرات b ومصفوفة القيود R وعددها g .

وبعد حساب F من العلاقة (5–48) أو(5–50) نقوم بمقارنة قيمتها المحسوبة مع القيمة الحرجة $F_{g,n-k}(\propto)$ وعند مستوى الدلالة ∝ . ونتخذ القرار حول معنوية التقديرات المقيدة $\tilde{\beta}_R$ كما يلي:

إذا كانت $F \le F_{g,n-k}(\propto)$ نقبل فرضية العدم H_0 (القيود صحيحة) .

وإذا كانت $F > F_{g,n-k}(\propto)$ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 التي تعني أن القيود غير صحيحة .

كما يمكن إنشاء مجال ثقة لكل من التقديرات المقيدة $\tilde{\beta}_R$. حسب القواعد المذكورة سابقاً للتقديرات غير المقيدة b .

مثال (5–2):

ادرس العلاقة بين الإنفاق على الاتصالات Y وعدد أفراد الأسرة X. وذلك بناء على بيانات عينة مؤلفة من $(n = 7)$ أسر والمبينة في الجدول التالي :

جدول (5–7): بيانات المسألة

المتوسط	7	6	5	4	3	2	1	رقم الأسرة
$\bar{X} = 5$	10	2	3	6	5	4	5	عدد الأفراد x_i
$\bar{Y} = 14$	17	11	13	14	12	16	15	مقدار الإنفاق بالآلاف Y

وذلك تحت القيود الخطية التالية (كل على حدة) .

$$\beta_0 + \beta_1 = 5 \qquad \text{ثم} \qquad \beta_0 = 10$$

الحل: نقوم أولاً بحساب التقديرات غير المقيدة b من العلاقة (1–9) أو من العلاقة التالية:

$$b = (X'X)^{-1} * X'Y$$

لذلك نشكل المصفوفة X الموسعة كما يلي:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad X' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

وبذلك نجد أن:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} * \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n, & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

وبعد الحساب نجد أن:

$$X'X = \begin{bmatrix} 7 & 35 \\ 35 & 215 \end{bmatrix}$$

ومنه نجد أن:

$$(X'X)^{-1} = \frac{1}{280}\begin{bmatrix} 215 & -35 \\ -35 & 7 \end{bmatrix}$$

وكذلك نجد أن:

$$X'Y = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_0 \\ \sum x_i y_i \end{bmatrix} = \begin{bmatrix} 98 \\ 514 \end{bmatrix}$$

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \frac{1}{280}\begin{bmatrix} 215 & -35 \\ -35 & 7 \end{bmatrix}\begin{bmatrix} 98 \\ 514 \end{bmatrix} = \begin{bmatrix} 11 \\ 0.6 \end{bmatrix}$$

ومنها نجد أن نموذج الانحدار المقدر غير المقيد يأخذ الشكل التالي :

$$\tilde{Y} = 11 + 0.6X$$

ولاختبار معنوية β_0 و β_1 نقوم بحساب المجاميع التالية:

$$SST = Y'Y - n\bar{Y}^2 = 1400 - 7(14)^2 = 28$$
$$SSX = b'X'Y - n\bar{Y}^2 = 1386.4 - 1372 = 14.4$$
$$SSE = SST - SSX = 28 - 14.4 = 13.6$$
$$\tilde{\sigma}_e^2 = \frac{SSE}{n-2} = \frac{13.6}{5} = 2.72$$
$$Var(b) = \tilde{\sigma}_e^2 (X'X)^{-1} = \frac{2.72}{280}\begin{bmatrix} 215 & -35 \\ -35 & 7 \end{bmatrix} = \begin{bmatrix} 2.089 & -0.345 \\ -0.345 & 0.068 \end{bmatrix}$$

أي أن:

$$Var(b_0) = 2.089 \qquad Var(b_1) = 0.068$$
$$\tilde{\sigma}_{b_0} = 1.445 \qquad \tilde{\sigma}_{b_1} = 0.261$$

وبناءً على هذه الحسابات نقوم باختبار معنوية β_0 و β_1 فنجد أن:

أولاً: لاختبار معنوية $\boldsymbol{\beta_0}$: نضع الفرضيتين كما يلي:

$$H_0 : \beta_0 = 0 \qquad , \qquad H_1 : \beta_0 \neq 0$$

ثم نقوم بحساب قيمة المؤشر t من العلاقة :

$$t = \frac{b_0 - (\beta_0)_0}{\tilde{\sigma}_{b_0}} = \frac{11 - 0}{1.445} = 7.612$$

ومن جداول t نجد أن القيمة الحرجة لـ t عندما $\propto = 0.5$ و $v = n - 2$ يساوي:

$$t_{n-2}\left(\frac{\propto}{2}\right) = t_5(0.025) = 2.571$$

وعند المقارنة نجد أن $t > t_5(\propto)$ ، لذلك نرفض فرضية العدم H_0 ونقبل الفرضية البديلة التي تفيد أن : H_1 $\beta_1 \neq 0$ أي أن قيمة β_0 معنوية .

ثانياً: لاختبار معنوية $\boldsymbol{\beta_1}$: نضع الفرضيتين كما يلي:

$$H_0 : \beta_1 = 0 \quad , \quad H_1 : \beta_1 \neq 0$$

ثم نقوم بحساب مؤشر الاختبار t من العلاقة :

$$t = \frac{b_1 - (\beta_1)_0}{\tilde{\sigma}_{b_1}} = \frac{0.6}{0.261} = 2.299$$

وبالمقارنة بالقيمة الحرجة السابقة: $t_{n-2}\left(\frac{\propto}{2}\right)$ التي تساوي أيضاً 2.571 نجد أن:

$t < t_5(\propto)$ ، لذلك نقبل فرضية العدم H_0 والتي تفيد أن قيمة β_1 غير معنوية لأن $H_0 : \beta_1 = 0$.

والآن نقوم بإعادة تقدير معالم النموذج المقيد β_0 وβ_1, تحت القيد الأول فقط وهو $\beta_0 = 10$، والذي يمكن كتابته كما يلي:

$$R_1\beta = C$$

$$[1 \;\; 0]\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} = 10 \;\; \Rightarrow \;\; \beta_0 - 0 = 10$$

ولحساب عناصر العلاقة (1−50) نقوم بحساب ما يلي:

$$(X'X)^{-1} * R_1' = \frac{1}{280}\begin{bmatrix}215 & -35\\ -35 & 7\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix} = \begin{bmatrix}\frac{215}{280}\\ \frac{-35}{280}\end{bmatrix}$$

ثم نقوم بحساب الجداء التالي :

$$R_1(X'X)^{-1}R_1' = [1 \;\; 0]\begin{bmatrix}\frac{215}{280}\\ \frac{-35}{280}\end{bmatrix} = \frac{215}{280} \qquad \text{(عدد)}$$

ثم نحسب المقلوب التالي :

$$[R_1(X'X)^{-1}R_1']^{-1} = \frac{280}{215}$$

وكذلك نحسب الحد التالي :

$$R_1b - C = [1 \;\; 0]\begin{bmatrix}11\\ 0.6\end{bmatrix} - 10 = 11 - 10 = 1$$

وبذلك نجد أن التقديرات المقيدة لـ β_R تساوي حسب (5−45) تساوي ما يلي:

$$\tilde{\beta}_{R_1} = \begin{bmatrix}11\\ 0.6\end{bmatrix} - \begin{bmatrix}\frac{215}{280}\\ \frac{-35}{280}\end{bmatrix}\left(\frac{280}{215}\right)(1) = \begin{bmatrix}11\\ 0.6\end{bmatrix} - \begin{bmatrix}1\\ -0.1627\end{bmatrix} = \begin{bmatrix}10\\ 0.7627\end{bmatrix}$$

وهنا نلاحظ أن القيد المفروض قد تحقق وأعطانا أن: $\tilde{\beta}_{oR_1} = 10$

ولكن قيمة β_1 أصبحت تساوي: 0.7267

وإن النموذج المقيد يأخذ الشكل التالي : $\tilde{X}_R = 10 + 0.7627X$

ولاختبار صحة القيد الأول نضع الفرضيتين كما يلي:

$$H_0 : R_1\beta = C \iff H_0 : \beta_0 = 10$$
$$H_1 : R_1\beta = C \iff H_1 : \beta_0 \neq 10$$

ثم نقوم بحساب قيمة مؤشر الاختبار F من العلاقة التالية:

$$F = \frac{(\boldsymbol{R_1 b - C})' \left[\boldsymbol{R}\left(\boldsymbol{X'X}\right)^{\boldsymbol{-1}}\boldsymbol{R_1'}\right]^{\boldsymbol{-1}} (\boldsymbol{R_1 b - C})/g}{\boldsymbol{ee'}/(n-2)} =$$

$$F = \frac{\left(1 * \frac{280}{215} * 1\right)/1}{2.72} = 0.4788$$

ثم نقوم بحساب القيمة الحرجة لـ F عندما $\propto = 0.05$ وعند درجتي الحرية $v_1 = g = 1$ و$v_2 = n - k = 5$ فنجد أن:

$$F_{g\ n-k}(\propto) = F_{1,5}(0.05) = 6.61$$

وعند المقارنة نجد أن: $F < F_{1,5}(\propto)$ ، لذلك نقبل فرضية العدم H_0 التي تقول أن القيمة $\beta_0 = 10$ هي قيمة معنوية وأن القيد الأول عليها هو قيد صحيح ومقبول .

والآن نقوم بإعادة تقدير معالم للنموذج المقيد بالشرط التالي:

$$\beta_0 + \beta_1 = 5$$

والذي يمكن كتابته على الشكل التالي :

$$R_2\beta = C \implies [1\ \ 1]\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} = 5$$

ثم نقوم بحساب العناصر التالية:

$$(X'X)^{-1}R_2' = \frac{1}{280}\begin{bmatrix}215 & -35\\ -35 & 7\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} = \begin{bmatrix}\frac{180}{280}\\ \frac{-28}{280}\end{bmatrix}$$

$$R_2(X'X)^{-1}R_2' = [1\ ,1]\begin{bmatrix}\frac{180}{280}\\ \frac{-28}{280}\end{bmatrix} = \frac{152}{280} \qquad (\text{عدد})$$

$$[R_2(X'X)^{-1}R_2']^{-1} = \frac{280}{152}$$

$$R_2 b - C = [1\ ,1]\begin{bmatrix}11\\ 0.6\end{bmatrix} - 5 = 11.6 - 5 = 6.6$$

وبذلك نجد أن التقديرات المقيدة بالشرط الثاني $\left(\beta_0 + \beta_1 = 5\right)$ تحسب من العلاقة(5−50) كما يلي:

$$\tilde{\beta}_{oR_2} = \begin{bmatrix}11\\ 0.6\end{bmatrix} - \begin{bmatrix}\frac{180}{280}\\ \frac{-28}{280}\end{bmatrix}\left(\frac{280}{152}\right)(6.6) = \begin{bmatrix}11\\ 0.6\end{bmatrix} - \begin{bmatrix}7.815\\ -1.215\end{bmatrix} = \begin{bmatrix}3.185\\ 1.815\end{bmatrix}$$

وهنا نلاحظ أن القيد الثاني قد تحقق وأصبح المجموع $\left(\beta_0 + \beta_1 = 5\right)$, ولاختبار صحة القيد الثاني نضع الفرضيتين على الشكل التالي :

$$H_0 : R_2\beta = C \iff H_0 : \beta_0 + \beta_1 = 5$$
$$H_1 : R_2\beta \neq C \iff H_1 : \beta_0 + \beta_1 \neq 5$$

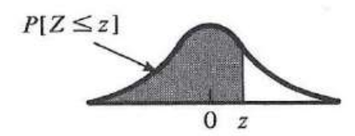
ثم نقوم بحساب قيمة مؤشر الاختبار من العلاقة (5−50) فنجد أن:

$$F = \frac{(6.6)\left(\frac{280}{152}\right)(6.6)/1}{2.72} = \frac{80.242}{2.72} = 29.500$$

وبمقارنة هذه القيمة لـ F مع القيمة الحرجة (6.61) نجد أن: $F > F_{1,5}(\propto)$ لذلك نرفض فرضية العدم H_0 ونقبل الفرضية H_1 التي تفيد أن القيد الثاني المفروض غير صحيح بالنسبة للمثال المدروس .

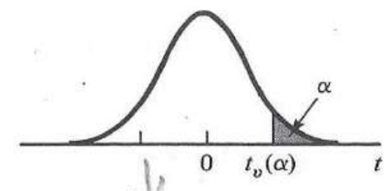
الملاحق: الجداول الاحصائية

TABLE 1 STANDARD NORMAL PROBABILITIES



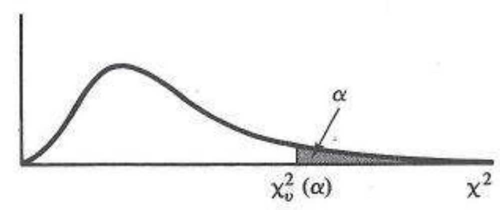
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

TABLE 2 STUDENT'S *t*-DISTRIBUTION CRITICAL POINTS



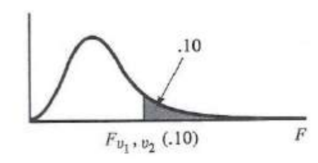
d.f. ν	.250	.100	.050	α .025	.010	.00833	.00625	.005
1	1.000	3.078	6.314	12.706	31.821	38.190	50.923	63.657
2	.816	1.886	2.920	4.303	6.965	7.649	8.860	9.925
3	.765	1.638	2.353	3.182	4.541	4.857	5.392	5.841
4	.741	1.533	2.132	2.776	3.747	3.961	4.315	4.604
5	.727	1.476	2.015	2.571	3.365	3.534	3.810	4.032
6	.718	1.440	1.943	2.447	3.143	3.287	3.521	3.707
7	.711	1.415	1.895	2.365	2.998	3.128	3.335	3.499
8	.706	1.397	1.860	2.306	2.896	3.016	3.206	3.355
9	.703	1.383	1.833	2.262	2.821	2.933	3.111	3.250
10	.700	1.372	1.812	2.228	2.764	2.870	3.038	3.169
11	.697	1.363	1.796	2.201	2.718	2.820	2.981	3.106
12	.695	1.356	1.782	2.179	2.681	2.779	2.934	3.055
13	.694	1.350	1.771	2.160	2.650	2.746	2.896	3.012
14	.692	1.345	1.761	2.145	2.624	2.718	2.864	2.977
15	.691	1.341	1.753	2.131	2.602	2.694	2.837	2.947
16	.690	1.337	1.746	2.120	2.583	2.673	2.813	2.921
17	.689	1.333	1.740	2.110	2.567	2.655	2.793	2.898
18	.688	1.330	1.734	2.101	2.552	2.639	2.775	2.878
19	.688	1.328	1.729	2.093	2.539	2.625	2.759	2.861
20	.687	1.325	1.725	2.086	2.528	2.613	2.744	2.845
21	.686	1.323	1.721	2.080	2.518	2.601	2.732	2.831
22	.686	1.321	1.717	2.074	2.508	2.591	2.720	2.819
23	.685	1.319	1.714	2.069	2.500	2.582	2.710	2.807
24	.685	1.318	1.711	2.064	2.492	2.574	2.700	2.797
25	.684	1.316	1.708	2.060	2.485	2.566	2.692	2.787
26	.684	1.315	1.706	2.056	2.479	2.559	2.684	2.779
27	.684	1.314	1.703	2.052	2.473	2.552	2.676	2.771
28	.683	1.313	1.701	2.048	2.467	2.546	2.669	2.763
29	.683	1.311	1.699	2.045	2.462	2.541	2.663	2.756
30	.683	1.310	1.697	2.042	2.457	2.536	2.657	2.750
40	.681	1.303	1.684	2.021	2.423	2.499	2.616	2.704
60	.679	1.296	1.671	2.000	2.390	2.463	2.575	2.660
120	.677	1.289	1.658	1.980	2.358	2.428	2.536	2.617
∞	.674	1.282	1.645	1.960	2.326	2.394	2.498	2.576

TABLE 3 χ^2 CRITICAL POINTS



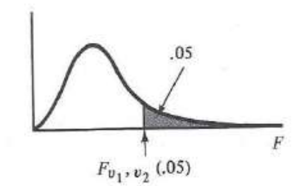
d.f. ν	.990	.950	.900	α .500	.100	.050	.025	.010	.005
1	.0002	.004	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.02	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.11	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.30	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.55	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.87	1.64	2.20	5.35	10.64	12.59	14.45	16.81	18.55
7	1.24	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.65	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.95
9	2.09	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.56	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	3.05	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.57	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	4.11	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.66	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	5.23	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.81	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	6.41	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	7.01	9.39	10.86	17.34	25.99	28.87	31.53	34.81	37.16
19	7.63	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	8.26	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.90	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	9.54	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	10.20	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	10.86	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	11.52	14.61	16.47	24.34	34.38	37.65	40.65	44.31	46.93
26	12.20	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	12.88	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.64
28	13.56	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	14.26	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	14.95	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	22.16	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	29.71	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	37.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	45.44	51.74	55.33	69.33	85.53	90.53	95.02	100.43	104.21
80	53.54	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	61.75	69.13	73.29	89.33	107.57	113.15	118.14	124.12	128.30
100	70.06	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

TABLE 4 F-DISTRIBUTION CRITICAL POINTS ($\alpha = .10$)



v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.05	62.26	62.53	62.79
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.17	5.17	5.16	5.15
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.81	2.80	2.78	2.76
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.57	2.56	2.54	2.51
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.27	2.25	2.23	2.21
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.17	2.16	2.13	2.11
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.03	2.01	1.99	1.96
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.93	1.91	1.89	1.86
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.89	1.87	1.85	1.82
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.86	1.84	1.81	1.78
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.83	1.81	1.78	1.75
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.80	1.78	1.75	1.72
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.78	1.76	1.73	1.70
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.76	1.74	1.71	1.68
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.74	1.72	1.69	1.66
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.71	1.69	1.66	1.62
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.68	1.66	1.63	1.59
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.67	1.65	1.61	1.58
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.66	1.64	1.60	1.57
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.65	1.63	1.59	1.56
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.64	1.62	1.58	1.55
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.63	1.61	1.57	1.54
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.50	1.48	1.44	1.40
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24

TABLE 5 *F*-DISTRIBUTION CRITICAL POINTS (α = .05)



v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.3	250.1	251.1	252.2
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.57
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.43
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.74
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.30
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.01
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.79
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.62
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.49
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.38
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.30
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.22
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.16
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.11
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.06
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.02
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.95
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.89
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.86
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.94	1.90	1.85	1.80
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.92	1.88	1.84	1.79
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.89	1.85	1.81	1.75
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.74
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.64
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.53
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.60	1.55	1.50	1.43
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.51	1.46	1.39	1.32

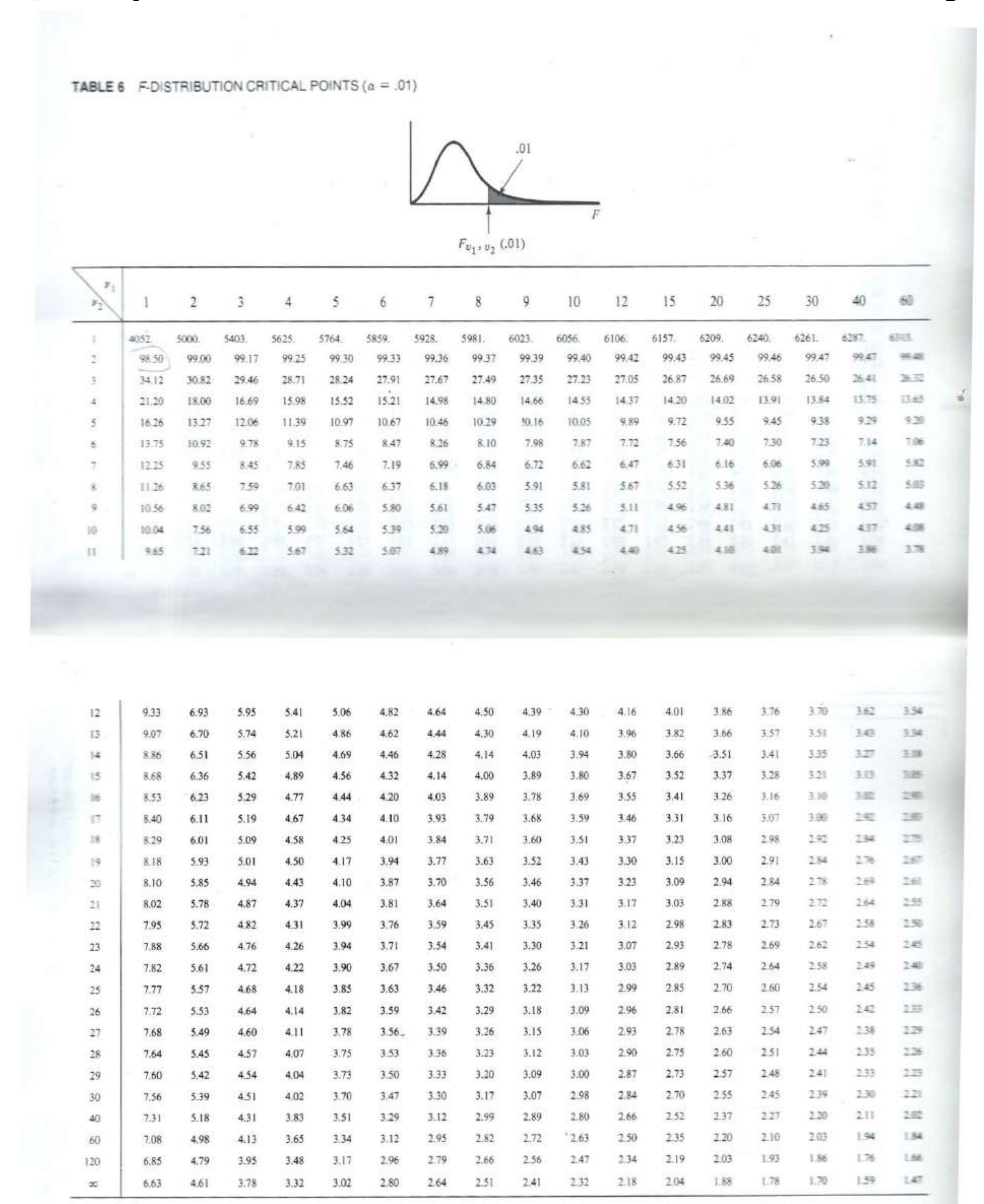


TABLE 6 F-DISTRIBUTION CRITICAL POINTS (α = .01)

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60
1	4052.	5000.	5403.	5625.	5764.	5859.	5928.	5981.	6023.	6056.	6106.	6157.	6209.	6240.	6261.	6287.	6313.
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.58	26.50	26.41	26.32
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.91	13.84	13.75	13.65
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.45	9.38	9.29	9.20
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.06
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.82
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.26	5.20	5.12	5.03
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.48
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.31	4.25	4.17	4.08
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.78
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.76	3.70	3.62	3.54
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.57	3.51	3.43	3.34
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.41	3.35	3.27	3.18
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.05
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.16	3.10	3.02	2.93
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.83
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	2.98	2.92	2.84	2.75
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.91	2.84	2.76	2.67
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.84	2.78	2.69	2.61
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.79	2.72	2.64	2.55
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.73	2.67	2.58	2.50
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.69	2.62	2.54	2.45
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.64	2.58	2.49	2.40
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.60	2.54	2.45	2.36
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.57	2.50	2.42	2.33
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.54	2.47	2.38	2.29
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.51	2.44	2.35	2.26
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.48	2.41	2.33	2.23
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.45	2.39	2.30	2.21
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.27	2.20	2.11	2.02
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.10	2.03	1.94	1.84
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.93	1.86	1.76	1.66
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.78	1.70	1.59	1.47